

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 19.3719 Seconds

(without alignments)
.1530.854 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MYPGANGGALAPARAGRL.....YSAPRDLTPPLNQTANTALL 715

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	362.5	9.6	925 1 NRP2_RAT	035276 rattus norv
2	354.5	9.4	931 1 NRP2_MOUSE	035375 mus musculu
3	351.5	9.3	931 1 NRP2_HUMAN	060462 homo sapien
4	338.5	9.0	923 1 NRP1_MOUSE	097333 mus musculu
5	338	8.9	914 1 NRP1_CHICK	019795 gallus gall
6	335	8.9	923 1 NRP1_HUMAN	014786 homo sapien
7	331	8.8	922 1 NRP1_RAT	099419 rattus norv
8	331	8.8	2211 1 FAS_BOVIN	028107 bos taurus
9	330	8.8	928 1 NRP1_XENLA	P28824 xenopus lae
10	317	8.4	2224 1 FAS_HUMAN	P12259 homo sapien
11	314	8.3	2258 1 FAS_PIG	099181 sus scrofa
12	299	7.9	387 1 MFGM_HUMAN	008431 mus musculu
13	297.5	7.9	427 1 MFGM_RAT	P70480 rattus norv
14	285.5	7.6	2351 1 FAS_HUMAN	P00451 homo sapien
15	276.5	7.3	463 1 MFGM_MOUSE	P21936 mus musculu
16	276.5	7.3	2319 1 FAS_MOUSE	006194 mus musculu
17	271.5	7.2	2133 1 FAS_PIG	P12283 sus scrofa
18	268	7.1	427 1 MFGM_BOVIN	095114 bos taurus
19	259.5	6.9	409 1 MFGM_PIG	P79385 sus scrofa
20	229.5	6.1	280 1 XLR1_FUGRU	099675 fugu rudrip
21	221.5	5.9	224 1 XLR1_MOUSE	092114 mus musculu
22	218.5	5.8	224 1 XLR1_HUMAN	015537 homo sapien
23	199	5.3	1308 1 CTAA_HUMAN	090800 homo sapien
24	195.5	5.2	1331 1 CTAA_HUMAN	090800 homo sapien
25	194.5	5.2	1310 1 CTAA_MOUSE	099467 mus musculu
26	192.5	5.1	1284 1 NRP4_DROME	094887 drosophila
27	190	5.0	3133 1 HMC7_HOMO	P98092 bombyx mori
28	186.5	4.9	666 1 MAS2_HUMAN	000187 homo sapien
29	186.5	4.9	1288 1 CTAA_HUMAN	092876 homo sapien
30	172.5	4.6	514 1 UVS2_XENLA	P42664 xenopus lae
31	165	4.4	986 1 BMP1_HUMAN	P13497 homo sapien
32	163	4.3	468 1 PCOI_RAT	008628 rattus norv
33	162.5	4.3	704 1 CRAR_MOUSE	P98064 mus musculu

34	162.5	4.3	854 1 DDR2_MOUSE	062371 mus musculu
35	161.5	4.3	449 1 PCOI_HUMAN	015113 homo sapien
36	160	4.2	468 1 PCOI_MOUSE	061398 mus musculu
37	159	4.2	550 1 COCH_HUMAN	043405 homo sapien
38	158	4.2	991 1 BMP1_MOUSE	P98063 mus musculu
39	158	4.2	1022 1 TLID_BRARE	057460 brachydanio
40	157	4.2	552 1 COCH_MOUSE	062507 mus musculu
41	156	4.1	707 1 BMP1_XENLA	P98070 xenopus lae
42	155.5	4.1	855 1 DDR2_HUMAN	016832 homo sapien
43	149.5	4.0	695 1 CASP_MESAU	P15156 mesocricetu
44	142.5	3.8	1385 1 CTAL_MOUSE	054991 mus musculu
45	142	3.8	855 1 ST14_HUMAN	097576 homo sapien

ALIGNMENTS

RESULT 1	ID	NRP2_RAT	STANDARD;	PRT;	925 AA.
AC	035276;				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).				
GN	NRP2.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Sprague-Dawley;				
RX	MEDLINE=97433085; Pubmed=9288754;				
RA	Kotodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,				
RA	Ginty D.D.;				
RT	"Neuropilin is a semaphorin III receptor."				
RL	Cell 90:753-762(1997).				
CC	-1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165				
CC	-1- AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.				
CC	-1- SUBCELLULAR LOCATION: TYPE I membrane protein.				
CC	-1- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE				
CC	CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE				
CC	LINING IN THE RIBS.				
CC	-1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.				
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 2 F3/8 TYPE C DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.				
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL: AF016297; AAC53338.1; -				
DR	HSSP: P12259; IC2T.				
DR	InterPro: IPR000859; CUB_domain.				
DR	InterPro: IPR000421; FAS5_C.				
DR	InterPro: IPR000998; MAM_domain.				
DR	Pfam: PF00431; CUB; 2.				
DR	Pfam: PF00629; MAM; 1.				
DR	Pfam: PF00754; F5_F8_Type_C; 2.				
DR	SMART: SM00242; CUB; 2.				
DR	SMART: SM00231; FAS5C; 2.				
DR	SMART: SM00137; MAM; 1.				
DR	PROSITE: PS01160; CUB; 2.				
DR	PROSITE: PS01285; FAS5C; 1; 2.				
DR	PROSITE: PS01286; FAS5C; 2; 2.				
DR	PROSITE: PS50060; MAM; 2; 1.				

Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.

FT SIGNAL 1 22
FT CHAIN 23 925
FT DOMAIN 23 858
FT TRANSMEM 859 883
FT DOMAIN 884 925
FT DOMAIN 28 142
FT DOMAIN 149 267
FT DOMAIN 277 427
FT DOMAIN 434 592
FT DOMAIN 642 802
FT DISULFID 83 105
FT DISULFID 149 175
FT DISULFID 208 230
FT DISULFID 277 427
FT DISULFID 434 592
FT CARBOHYD 152 152
FT CARBOHYD 157 157
FT CARBOHYD 629 629
FT CARBOHYD 833 833
FT CARBOHYD 834 834
SQ SEQUENCE 925 AA; 103896 MW; 3BF62903F644851C CRC64;

Query Match 9.6%; Score 362.5; DB 1; Length 925.
Best Local Similarity 22.3%; Pred. No. 7.2e-18;
Matches 143; Conservative 100; Mismatches 246; Indels 151; Gaps 24;

41 CGHLYTODSGTMTSKNPGTYPNHTVCEKITVPR-KGRLLRLRG-DIDIESQTCASDY 98
28 CGGRINSDAGYITSPGYRQDYPHONCEWVYAPENPKIYLNPNPHEIEKHOCKYDF 87
99 LIF-----TSSDQYGPYCGSMTPKELLINSEVTVRESGSHISGRGLTY-----AS 149
88 IEIRDGESEADLLKHCIGNI-APPTIISGSLYIKFSDYAROGAGFSLREIFKTS 146
150 SD-----HPLIYC-----LRASHYLTREYSK----- 173
147 EDCSNFTSPNGTISPGEPPEKYPHNLCTFTILAKPMEILLPLTDLHDPLOYGEG 206
174 -CPACRDVAGDIS--GNMVDGYRDTSLCKAAIHAGIADLGGQISVLRKGISRY-- 228
207 DCKYMWLDIMGSIHVGGLICKYCTKTPSKRSSTGISLTFPHDMVAVKCFARYYL 266
229 -----EGLANGVLSRD-GSLSDKRFLETSNGCSRSLSFEPPDQGITRASSSMQVNESGD 281
267 VHOEPPEFQCAAPLGMESGRIANE-----QISASSTP-----S 300
282 QVHNSPGARLADOGSPNASGSSNNHNPREFLEIDLGKKKITGIRTTG--STQSNFNF 339
301 DGRWTPQOSRLHGDNGMTPNVDSN---KEYLDVLRFLMLTALATQGAISRETQKGY 356
340 YKSFVAFNFKNNNSKMYKTYGIVNNEKVFQGSNFRDPVQNNFTPPYARVRYVPQTM 399
357 YKSYKLELVSTNGEDMWYRHKGNH--KYFQANNDATELVYLNKMLPLTLTRIRIRPQTM 414
400 HORILAKVELLGCQITQGNDSLVMKTS---OSTSVSTKKEDETTTPT----- 445
415 HLGIALRLLELCRCRYTDAPCSNNMLGLJADTQISASSTREYLMSPSAARLVSRSRG 474
446 -----PSEETSGINNTTVAIPVLVLYLVFAGMGIFA---AFPRKKKKGSPYSA 493
475 FPRPDQAPGEEW-LQVLDIGTPTKYGVYIQAGRGDSITLAEARAFYKKEFVVSISLNGK 533
494 EAOKTDCWKQIKYFPAHQSA-EFTISYD-----NEKEMTOKL 530
534 D-----WEYIQDPTQPKLEEGNMHYDTPDIRRFPVPYPAQYVYVERMSPAIGARL 587
531 DLITSDMADYQOPL-MIGTGIVTRKGSFFRPMDDAEAG 569
588 EYLGCDMTDSKPYETGLPTVYKSEETTPPYPMDEDATECG 627

RESULT 2
NRP2_MOUSE
ID NR2_MOUSE STANDARD; PRI; 931 AA.
AC 035375; 035373; 035374; 035376; 035377; 035378;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neutropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
GN NR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
[1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN-BALB/c;
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neutropilin-2, a novel member of the neutropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
CC - FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOPRFORMS OF VEGF, AND THE PLGF-2 ISOPRFORM OF PGF.
CC - SUBUNIT: NEUTROPILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUTROPILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - ALTERNATIVE PROLOCATIONS: 6 ISOFORMS; A22 (SHOWN HERE), A0, A5, A17,
CC B0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC - TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
CC NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,
CC INTESTINAL EPITHELIUM, KIDNEY, LONG AND SUBMANDIBULAR GLAND.
CC - DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
CC IS DEVELOPMENTALLY REGULATED.
CC - SIMILARITY: BELONGS TO THE NEUTROPILIN FAMILY.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; AF022856; AAC53379.1; -
DR EMBL; AF022854; AAC53377.1; -
DR EMBL; AF022855; AAC53378.1; -
DR EMBL; AF022857; AAC53380.1; -
DR EMBL; AF022858; AAC53381.1; -
DR EMBL; AF022861; AAC53382.1; -
DR HSSP; P12259; 1CCT.
DR MGD; MGI:1100492; NRP2.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS50060; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 931
FT POTENTIAL.
FT NEUTROPILIN-2.

DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00231; FA58C; 2.
 DR SMART: SM00137; MAM; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01285; FA58C; 1; 2.
 DR PROSITE: PS01286; FA58C; 2; 2.
 DR PROSITE: PS00740; MAM; 1; 1.
 DR PROSITE: PS50060; MAM; 2; 1.
 DR Transmembrane: Glycoprotein; Neurone; Signal; Repeat; Receptor.
 FT SIGNAL 1 21
 FT CHAIN 1 923
 FT DOMAIN 22 856
 FT TRANSMEM 857 879
 FT DOMAIN 880 923
 FT DOMAIN 27 141
 FT DOMAIN 147 265
 FT DOMAIN 275 424
 FT DOMAIN 431 583
 FT DOMAIN 645 811
 FT DISULFID 27 54
 FT DISULFID 82 104
 FT DISULFID 147 173
 FT DISULFID 206 228
 FT DISULFID 275 424
 FT DISULFID 431 583
 FT CARBOHYD 150 261
 FT CARBOHYD 261 300
 FT CARBOHYD 300 322
 FT CARBOHYD 522 583
 FT CARBOHYD 842 842
 SQ SEQUENCE 923 AA; 103020 MW; 06448BA170796808 CRC64;

Query Match 9.0%; Score 338.5; DB 1; Length 923;

Best Local Similarity 20.8%; Pred. No. 3.6e-16; Matches 158; Conservative 106; Mismatches 267; Indels 227; Gaps 30;

DR 17 RGL-----LALLAVSAPLRQLQAEELGDCGHLVYODSGTMSKNYPGTYPNHTYCEK 70
 DB 3 RGLPLLCATLALALALAGNR-----SDKCGGTIKIENGNYLSTPGYRHSYSEKCEW 56
 QY 71 TTPVPG-KRLIIRLG-DIDIESQTCASDYLFTSSDQ---YGPQSGMTVPKELLN 124
 DB 57 LQAPPEYRIIINENPHRDLEDRCKYVEVIDENEGRLMKFCGKI-APSPVSS 115
 QY 125 TSEVTVRPFSGSHISGRFLTYA-----SSDHDLLTC-- 158
 DB 116 GPLLFTKEVSDYETHGAGFSIRIETIKRQPECSQNTAPTGVIKSGPPEKYNCLCTY 175
 QY 159 -----LERASHYKTEYSKPCPAG--CRVAGDISGNMVDGYRDTSLCKRAIH 205
 DB 176 IIFAPKSEIILEFESFDLEQDSNP--PGMPCRYDRLEI--WQGFPE-----VGRH 223
 QY 206 AG-TINDELGGOT-----SVLRKGISRYEGITLANGVLSRDGSLSDRFLTE 250
 DB 224 IGYCCQKPKPRIRSSSGVLSMVFYTDLSAIKGFSGANSVLQSSI-SEDFK----- 274
 QY 251 TSNCGSRISLSEF-----DQDIRASSSMQSVNESCQDVHHSFGQARLQDQPSMASGSSN 306
 DB 275 -----CHEALGMSGEHSDQITNSQYGT-----MNSVESRSLNTPENGMTPEEDST 322
 QY 307 NHRPREMLIDLGKKRKTGIRTTG--STQSNENFYVSKSFVNFKNNNSKMTYKGIYNN 364
 DB 323 ---KEMIQVDGLRFLVAVGTGAIKSKETKKYVVKYIRVDISSNGEDMISLK--EGN 376
 QY 365 EEKVEFGNSNFRDPVONNTPPIPVARYRVYVQTHQRIALAKVELIGCOITQGNDSLWVR 424
 DB 377 KAIIFQGNNTNPDIYVGVFSKPLITFRVRIKIPVSMETGISMREYVGCKITD-----YP 430
 QY 425 KTSQSSTVSSTKKREDEITRPIPESETSTGINITY-----AIPLVLVAVYVAGGIFA 478
 DB 431 CSGMLGMSGLISDSQITNSQADNRMMFENILVTSRTGALP----- 474

QY 479 AFRKKKKSGPYGSAEAKTDCWKQI-----KYPPARHQAFTTISYDN 522
 DB 475 -----PSRHPY-----INELQVLDGDEKIVRGVILIOGCHHRENNKPKFKIAYSN 521
 QY 523 EKEMTOKLIDITSDMADYQOPLMIGTGYTRKSGTFRPMDTAEAGVSTAGHYDCPO 582
 DB 522 -----NSDMWTIMDSKRRKAKSFGEGNNNTDTP 550
 QY 583 RAGRHEALPLAPPEPYATPI---VERHVLRAHTSAGSGIVNPPQPOPHKHSLSGGR 639
 DB 551 LRTFSPSTREIRIYPERATHSGIGLRLMELGCEVAPAPPTPMPGNPVHE----- 602
 QY 640 SPVAGVGAODPDYORPSPADRGYDRPRAVALATE 677
 DB 603 -----CDDDQANCHSGTGD--FOLGGTIVLATE 630

RESULT 5
 NRPL_CHICK
 ID NRPL_CHICK STANDARD; PRT; 914 AA.
 AC P79795;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilin-1 precursor (A5 protein).
 GN NRPL OR NRP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White Leghorn; TISSUE=Embryonic brain;
 RX MEDLINE=95324761; PubMed=7601310;
 RA Takagi S., Kasuya Y., Shimizu M., Matsuura T., Tsuboi M., Kawakami A.,
 RA Fujisawa H.;
 RT "Expression of a cell adhesion molecule, neurofilin, in the
 RT developing chick nervous system.";
 RT Dev. Biol. 170:207-222(1995).
 RL dev. Biol. 170:207-222(1995).
 CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANORENIN ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION
 CC PROPERTIES.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
 CC (LAYERS D AND E OF SGEs), AMACRINE CELLS OF RETINA, NEURITES OF
 CC DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
 CC BLOOD VESSELS IN THE ENTIRE EMBRYO.
 CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: D45416; BAA08256.1; -
 CC HSSP: P12259; ICZP.
 CC InterPro: IPR000859; CUB_domain.
 CC InterPro: IPR000421; FA58_C.
 CC InterPro: IPR000998; MAM_domain.
 CC Pfam: PF00431; CUB; 2.
 CC Pfam: PF00629; MAM; 1.
 CC Pfam: PF00754; F5_F8_type_C; 2.

FT DOMAIN 348 686 F5/8 TYPE A 2.
 FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 535 686 PLASTOCYANIN-LIKE 4.
 FT SIMILAR 696 1564 B.
 FT DOMAIN 899 915 TO 17 AA REPEATS IN HUMAN FAS.
 FT DOMAIN 1124 1151 2 X 14 AA TANDEM REPEATS.
 FT REPEAT 1124 1137 1-1.
 FT REPEAT 1124 1138 1-2.
 FT DOMAIN 1188 1453 30 X 9 AA APPROXIMATE TANDEM REPEATS OF
 [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
 FT REPEAT 1188 1196 2-1.
 FT REPEAT 1197 1205 2-2.
 FT REPEAT 1206 1214 2-3.
 FT REPEAT 1215 1223 2-4.
 FT REPEAT 1224 1232 2-5.
 FT REPEAT 1233 1241 2-6.
 FT REPEAT 1242 1250 2-7.
 FT REPEAT 1251 1259 2-8.
 FT REPEAT 1260 1268 2-9.
 FT REPEAT 1269 1277 2-10.
 FT REPEAT 1278 1286 2-11.
 FT REPEAT 1287 1295 2-12.
 FT REPEAT 1296 1304 2-13.
 FT REPEAT 1305 1313 2-14.
 FT REPEAT 1314 1322 2-15.
 FT REPEAT 1323 1331 2-16.
 FT REPEAT 1332 1340 2-17.
 FT REPEAT 1341 1349 2-18.
 FT REPEAT 1350 1358 2-19.
 FT REPEAT 1359 1367 2-20.
 FT REPEAT 1368 1376 2-21.
 FT REPEAT 1377 1385 2-22.
 FT REPEAT 1386 1394 2-23.
 FT REPEAT 1395 1403 2-24.
 FT REPEAT 1404 1412 2-25.
 FT REPEAT 1413 1421 2-26.
 FT REPEAT 1422 1430 2-27.
 FT REPEAT 1431 1439 2-28.
 FT REPEAT 1440 1444 2-29 (PARTIAL).
 FT REPEAT 1445 1453 2-30.
 FT DOMAIN 1569 1890 F5/8 TYPE A 3.
 FT DOMAIN 1569 1738 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 1894 2048 F5/8 TYPE C 1.
 FT DOMAIN 2053 2208 F5/8 TYPE C 2.
 FT SITE 741 742 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1564 1565 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT DISULFID 167 193 PROBABLE.
 FT DISULFID 499 525 PROBABLE.
 FT DISULFID 1712 1738 BY SIMILARITY.
 FT DISULFID 1894 2048 BY SIMILARITY.
 FT DISULFID 2053 2208 BY SIMILARITY.
 FT MOD_RES 697 701 SUPFATION (POTENTIAL).
 FT MOD_RES 701 701 SUPFATION (POTENTIAL).
 FT MOD_RES 730 730 SUPFATION (POTENTIAL).
 FT MOD_RES 1513 1513 SUPFATION (POTENTIAL).
 FT MOD_RES 1529 1529 SUPFATION (POTENTIAL).
 FT MOD_RES 1537 1537 SUPFATION (POTENTIAL).
 FT MOD_RES 1541 1541 SUPFATION (POTENTIAL).
 FT MOD_RES 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 587 592 N-TERMINAL (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 2211 AA; 248981 MW; CBBP90B738667C45 CRC64;
 Query Match 8.88; Score 331; DB 1; Length 2211;
 Best Local Similarity 32.58; Pred. No. 4,4e-15;
 Matches 88; Conservative 48; Mismatches 89; Indels 46; Gaps 10;
 QY 162 ASHYKTEY-SKCPA-----GCRDVGADISGNV--DGYRDSILCKAIIAGITADE 212
 DB 1968 AKHYLKPYTTEFCVAISLDRKNMRFKGNSTRVMTFGNSDASTIKENOIDPPVVA-- 2025
 QY 213 LGQISVLRKRGISREGLIANGVLSRDSLSDRKFLFTSGCSRSLSFE---PDGQIR 268
 DB 2026 -----RYRISPTGSKYKRALBLEQGC-EVAGCSTPLMGEGKIKNKQIT 2070
 QY 269 ASS---SWQSVNESGDQVWSPGQARLDQD--PSMAGSDSSNNHKPREMLDILGKERK 323
 DB 2071 ASSEKSKSWWG-----NYMEPLRLAAGRVANAKANNNN---QWQIDILKIKK 2119
 QY 324 ITGRTGSGTQSNFNYFKSVNMFKNNSKTKYTKGVNNEEVPOGNSFRPVQONF 383
 DB 2120 ITAIVTQCKSLSEMYKSTYTHSDGTMDKRYREKSSWYDKIFGNNNVKGVKNKF 2179
 QY 384 IPIVARYRVVPOTWRIALKEYLGCQI 414
 DB 2180 NPILSRIRIRIPRTWNSIALRLFCGCM 2210
 RESULT 9
 NRPL_XENIA STANDARD; PRT; 928 AA.
 AC P28824;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 24, Last annotation update)
 DE Neuropilin-1 precursor (A5 protein) (A5 antigen).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91337458; PubMed=1908252;
 RA Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;
 RT "The A5 antigen, a candidate for the neuronal recognition molecule,
 has homologues to complement components and coagulation factors.";
 RL Neuron 7:295-307(1991).
 CC - FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
 BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
 CC - SURCELLULAR LOCATION: TYPE I membrane protein.
 CC - TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
 CC NEURONS.
 CC - SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.

```

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D10467; BAA01260.1; .
CC HSSP; P12259; 1C2T.
CC InterPro: IPR000859; CUB_domain.
CC InterPro: IPR000421; FA58_C.
CC InterPro: IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC PRINTS; PR00020; MAMDOMAIN.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00231; FA58C; 2.
CC SMART; SM00137; MAM; 1.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS00600; MAM_2; 1.
CC Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Antigen.
KM SIGNAL.
FT 1 21 POTENTIAL.
FT 2 928 NEUTROFILIN-1.
FT 22 860 EXTRACELLULAR (POTENTIAL).
FT 22 860 TRANSMEM.
FT 861 883 POTENTIAL.
FT 884 928 CYTOPLASMIC (POTENTIAL).
FT 27 141 CUB 1.
FT 27 141 CUB 2.
FT 265 265 F5/8 TYPE C 1.
FT 424 424 F5/8 TYPE C 2.
FT 431 584 MAM.
FT 431 584 MAM.
FT 466 812 PROBABLE.
FT 27 54 PROBABLE.
FT 82 104 PROBABLE.
FT 147 173 PROBABLE.
FT 206 228 PROBABLE.
FT 275 424 BY SIMILARITY.
FT 431 584 BY SIMILARITY.
FT 431 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 928 AA; 103416 MW; AF6B323B0AAC789D CRC64;

Query Match 8.8%; Score 330; DB 1; Length 928;
Best Local Similarity 24.2%; Pred. NO. 1.5e-15;
Matches 112; Conservative 74; Mismatches 17; Indels 106; Gaps 20;

```

```

OY 213 LGGQISVLQKRGISRYEGILA-----NGVLSRDG-----SLSDKRFLETSNGCSRS 258
DB 224 IGRYCGMTPGVRNRFETGLSMIFHTDAISAKKEGFANFSVQSTDDF-----QCKEA 278
OY 259 LSEFPD---GOIRASSWQSVNESGDVHMSPGARLQDGPASASDSSNNKPREML 314
DB 279 LGMESGEIHFDDISVSQY-----SMNWSAERSRLNLYVENCWTPEEDT-----VKEMI 326
OY 315 EIDLKERRKIGIRFTG---STQSNFNFYKSPVMAEFKNNNSWKTKGIVNNEEKVPOGN 372
DB 327 QVDLENLRFVSGTQGAISKETKKYKTKVDISSNGEDWTLKO--GNKHLVFQGN 384
OY 373 SNFRDPVQNNFPIVARYVRVVPOTWQRIALKEVLCQQT 415
DB 385 TDATDVYVRPFESKPIYTRFVRRLRPYTWENGISLRELYGCKRT 427

RESULT 10
FAS_HUMAN
ID FAS_HUMAN STANDARD; PRT; 2224 AA.
AC P12259; O14285;
DT 01-OCT-1989 (Rel. 12, Created)
DI 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=92232668; PubMed=1567832;
RA Crilpe L.D., Moore K.D., Kane W.H.;
RT "Structure of the gene for human coagulation factor V.";
RL Biochemistry 31:3777-3785(1992).
[2]
SEQUENCE FROM N.A.
MEDLINE=87260886; PubMed=3110773;
RA Jenny R.J., Pittman D.D., Toole J.J., Kitz R.W., Aldape R.A.,
RA Hewick R.M., Kaufman R.J., Mann K.G.;
RT "Complete cDNA and derived amino acid sequence of human factor V.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
[3]
SEQUENCE OF 1-1600 FROM N.A.
MEDLINE=88107560; PubMed=2827731;
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
RT region of human factor V, a blood coagulation factor with four types
RT of internal repeats.";
RL Biochemistry 26:6508-6514(1987).
[4]
SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
MEDLINE=86313665; PubMed=3092220;
RA Kane W.H., Davie E.W.;
RT "Cloning of a cDNA coding for human factor V, a blood coagulation
RT factor homologous to factor VIII and ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
[5]
PARTIAL SEQUENCE FROM N.A.
TISUE=FBIDblast;
MEDLINE=93203619; PubMed=8454869;
RA Shen N.L., Fan S.-T., Pyatt J., Graff R., Lapolla R.J.,
RA Edgington T.S.;
RT "The serine protease cofactor factor V is synthesized by
RT lymphocytes.";
RL J. Immunol. 150:2992-3001(1993).
[6]
SULFATION.
MEDLINE=94264012; PubMed=8204629;
RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
RA Kaufman R.J.;

```

RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity."
 RL Biochemistry 33:6952-6959(1994).
 RN [17]
 RP SULFATION.
 RX MEDLINE-90366699; PubMed-2168225;
 RA Horton G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V."
 RL Blood 76:946-952(1990).
 RN [18]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE-20052169; PubMed-10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V."
 RL Nature 402:434-439(1999).
 RN [19]
 RP VARIANT APCR GLN-534.
 RX MEDLINE-94217810; PubMed-8164741;
 RA Bertina R.M., Koelmaan B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C."
 RL Nature 369:64-67(1994).
 RN [10]
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese."
 RL Blood 91:1135-1139(1998).
 RN [11]
 RP VARIANT APCR THR-334.
 RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C."
 RL Blood 91:1140-1144(1998).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 CC REPEATS.
 CC -1- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
 CC -1- DISEASE: OMEN PARAHEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
 CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
 CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
 CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
 CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
 CC IMPLANTATION.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; L32759; AAB59401.1; JOINED.
 DR EMBL; L32760; AAB59401.1; JOINED.
 DR EMBL; L32761; AAB59401.1; JOINED.
 DR EMBL; L32762; AAB59401.1; JOINED.
 DR EMBL; L32763; AAB59401.1; JOINED.
 DR EMBL; L32764; AAB59401.1; JOINED.
 DR EMBL; L32765; AAB59401.1; JOINED.
 DR EMBL; L32766; AAB59401.1; JOINED.
 DR EMBL; L32767; AAB59401.1; JOINED.
 DR EMBL; L32768; AAB59401.1; JOINED.
 DR EMBL; L32769; AAB59401.1; JOINED.
 DR EMBL; L32770; AAB59401.1; JOINED.
 DR EMBL; L32771; AAB59401.1; JOINED.
 DR EMBL; L32772; AAB59401.1; JOINED.
 DR EMBL; L32773; AAB59401.1; JOINED.
 DR EMBL; L32774; AAB59401.1; JOINED.
 DR EMBL; L32775; AAB59401.1; JOINED.
 DR EMBL; L32776; AAB59401.1; JOINED.
 DR EMBL; L32777; AAB59401.1; JOINED.
 DR EMBL; L32778; AAB59401.1; JOINED.
 DR EMBL; L32779; AAB59401.1; JOINED.
 DR EMBL; L32780; AAB59401.1; JOINED.
 DR EMBL; L32781; AAB59401.1; JOINED.
 DR EMBL; L32782; AAB59401.1; JOINED.
 DR EMBL; L32783; AAB59401.1; JOINED.
 DR EMBL; L32784; AAB59401.1; JOINED.
 DR EMBL; L32785; AAB59401.1; JOINED.
 DR EMBL; L32786; AAB59401.1; JOINED.
 DR EMBL; L32787; AAB59401.1; JOINED.
 DR EMBL; L32788; AAB59401.1; JOINED.
 DR EMBL; L32789; AAB59401.1; JOINED.
 DR EMBL; L32790; AAB59401.1; JOINED.
 DR EMBL; L32791; AAB59401.1; JOINED.
 DR EMBL; L32792; AAB59401.1; JOINED.
 DR EMBL; L32793; AAB59401.1; JOINED.
 DR EMBL; L32794; AAB59401.1; JOINED.
 DR EMBL; L32795; AAB59401.1; JOINED.
 DR EMBL; L32796; AAB59401.1; JOINED.
 DR EMBL; L32797; AAB59401.1; JOINED.
 DR EMBL; L32798; AAB59401.1; JOINED.
 DR EMBL; L32799; AAB59401.1; JOINED.
 DR EMBL; L32800; AAB59401.1; JOINED.
 DR EMBL; L32801; AAB59401.1; JOINED.
 DR EMBL; L32802; AAB59401.1; JOINED.
 DR EMBL; L32803; AAB59401.1; JOINED.
 DR EMBL; L32804; AAB59401.1; JOINED.
 DR EMBL; L32805; AAB59401.1; JOINED.
 DR EMBL; L32806; AAB59401.1; JOINED.
 DR EMBL; L32807; AAB59401.1; JOINED.
 DR EMBL; L32808; AAB59401.1; JOINED.
 DR EMBL; L32809; AAB59401.1; JOINED.
 DR EMBL; L32810; AAB59401.1; JOINED.
 DR EMBL; L32811; AAB59401.1; JOINED.
 DR EMBL; L32812; AAB59401.1; JOINED.
 DR EMBL; L32813; AAB59401.1; JOINED.
 DR EMBL; L32814; AAB59401.1; JOINED.
 DR EMBL; L32815; AAB59401.1; JOINED.
 DR EMBL; L32816; AAB59401.1; JOINED.
 DR EMBL; L32817; AAB59401.1; JOINED.
 DR EMBL; L32818; AAB59401.1; JOINED.
 DR EMBL; L32819; AAB59401.1; JOINED.
 DR EMBL; L32820; AAB59401.1; JOINED.
 DR EMBL; L32821; AAB59401.1; JOINED.
 DR EMBL; L32822; AAB59401.1; JOINED.
 DR EMBL; L32823; AAB59401.1; JOINED.
 DR EMBL; L32824; AAB59401.1; JOINED.
 DR EMBL; L32825; AAB59401.1; JOINED.
 DR EMBL; L32826; AAB59401.1; JOINED.
 DR EMBL; L32827; AAB59401.1; JOINED.
 DR EMBL; L32828; AAB59401.1; JOINED.
 DR EMBL; L32829; AAB59401.1; JOINED.
 DR EMBL; L32830; AAB59401.1; JOINED.
 DR EMBL; L32831; AAB59401.1; JOINED.
 DR EMBL; L32832; AAB59401.1; JOINED.
 DR EMBL; L32833; AAB59401.1; JOINED.
 DR EMBL; L32834; AAB59401.1; JOINED.
 DR EMBL; L32835; AAB59401.1; JOINED.
 DR EMBL; L32836; AAB59401.1; JOINED.
 DR EMBL; L32837; AAB59401.1; JOINED.
 DR EMBL; L32838; AAB59401.1; JOINED.
 DR EMBL; L32839; AAB59401.1; JOINED.
 DR EMBL; L32840; AAB59401.1; JOINED.
 DR EMBL; L32841; AAB59401.1; JOINED.
 DR EMBL; L32842; AAB59401.1; JOINED.
 DR EMBL; L32843; AAB59401.1; JOINED.
 DR EMBL; L32844; AAB59401.1; JOINED.
 DR EMBL; L32845; AAB59401.1; JOINED.
 DR EMBL; L32846; AAB59401.1; JOINED.
 DR EMBL; L32847; AAB59401.1; JOINED.
 DR EMBL; L32848; AAB59401.1; JOINED.
 DR EMBL; L32849; AAB59401.1; JOINED.
 DR EMBL; L32850; AAB59401.1; JOINED.
 DR EMBL; L32851; AAB59401.1; JOINED.
 DR EMBL; L32852; AAB59401.1; JOINED.
 DR EMBL; L32853; AAB59401.1; JOINED.
 DR EMBL; L32854; AAB59401.1; JOINED.
 DR EMBL; L32855; AAB59401.1; JOINED.
 DR EMBL; L32856; AAB59401.1; JOINED.
 DR EMBL; L32857; AAB59401.1; JOINED.
 DR EMBL; L32858; AAB59401.1; JOINED.
 DR EMBL; L32859; AAB59401.1; JOINED.
 DR EMBL; L32860; AAB59401.1; JOINED.
 DR EMBL; L32861; AAB59401.1; JOINED.
 DR EMBL; L32862; AAB59401.1; JOINED.
 DR EMBL; L32863; AAB59401.1; JOINED.
 DR EMBL; L32864; AAB59401.1; JOINED.
 DR EMBL; L32865; AAB59401.1; JOINED.
 DR EMBL; L32866; AAB59401.1; JOINED.
 DR EMBL; L32867; AAB59401.1; JOINED.
 DR EMBL; L32868; AAB59401.1; JOINED.
 DR EMBL; L32869; AAB59401.1; JOINED.
 DR EMBL; L32870; AAB59401.1; JOINED.
 DR EMBL; L32871; AAB59401.1; JOINED.
 DR EMBL; L32872; AAB59401.1; JOINED.
 DR EMBL; L32873; AAB59401.1; JOINED.
 DR EMBL; L32874; AAB59401.1; JOINED.
 DR EMBL; L32875; AAB59401.1; JOINED.
 DR EMBL; L32876; AAB59401.1; JOINED.
 DR EMBL; L32877; AAB59401.1; JOINED.
 DR EMBL; L32878; AAB59401.1; JOINED.
 DR EMBL; L32879; AAB59401.1; JOINED.
 DR EMBL; L32880; AAB59401.1; JOINED.
 DR EMBL; L32881; AAB59401.1; JOINED.
 DR EMBL; L32882; AAB59401.1; JOINED.
 DR EMBL; L32883; AAB59401.1; JOINED.
 DR EMBL; L32884; AAB59401.1; JOINED.
 DR EMBL; L32885; AAB59401.1; JOINED.
 DR EMBL; L32886; AAB59401.1; JOINED.
 DR EMBL; L32887; AAB59401.1; JOINED.
 DR EMBL; L32888; AAB59401.1; JOINED.
 DR EMBL; L32889; AAB59401.1; JOINED.
 DR EMBL; L32890; AAB59401.1; JOINED.
 DR EMBL; L32891; AAB59401.1; JOINED.
 DR EMBL; L32892; AAB59401.1; JOINED.
 DR EMBL; L32893; AAB59401.1; JOINED.
 DR EMBL; L32894; AAB59401.1; JOINED.
 DR EMBL; L32895; AAB59401.1; JOINED.
 DR EMBL; L32896; AAB59401.1; JOINED.
 DR EMBL; L32897; AAB59401.1; JOINED.
 DR EMBL; L32898; AAB59401.1; JOINED.
 DR EMBL; L32899; AAB59401.1; JOINED.
 DR EMBL; L32900; AAB59401.1; JOINED.
 DR EMBL; L32901; AAB59401.1; JOINED.
 DR EMBL; L32902; AAB59401.1; JOINED.
 DR EMBL; L32903; AAB59401.1; JOINED.
 DR EMBL; L32904; AAB59401.1; JOINED.
 DR EMBL; L32905; AAB59401.1; JOINED.
 DR EMBL; L32906; AAB59401.1; JOINED.
 DR EMBL; L32907; AAB59401.1; JOINED.
 DR EMBL; L32908; AAB59401.1; JOINED.
 DR EMBL; L32909; AAB59401.1; JOINED.
 DR EMBL; L32910; AAB59401.1; JOINED.
 DR EMBL; L32911; AAB59401.1; JOINED.
 DR EMBL; L32912; AAB59401.1; JOINED.
 DR EMBL; L32913; AAB59401.1; JOINED.
 DR EMBL; L32914; AAB59401.1; JOINED.
 DR EMBL; L32915; AAB59401.1; JOINED.
 DR EMBL; L32916; AAB59401.1; JOINED.
 DR EMBL; L32917; AAB59401.1; JOINED.
 DR EMBL; L32918; AAB59401.1; JOINED.
 DR EMBL; L32919; AAB59401.1; JOINED.
 DR EMBL; L32920; AAB59401.1; JOINED.
 DR EMBL; L32921; AAB59401.1; JOINED.
 DR EMBL; L32922; AAB59401.1; JOINED.
 DR EMBL; L32923; AAB59401.1; JOINED.
 DR EMBL; L32924; AAB59401.1; JOINED.
 DR EMBL; L32925; AAB59401.1; JOINED.
 DR EMBL; L32926; AAB59401.1; JOINED.
 DR EMBL; L32927; AAB59401.1; JOINED.
 DR EMBL; L32928; AAB59401.1; JOINED.
 DR EMBL; L32929; AAB59401.1; JOINED.
 DR EMBL; L32930; AAB59401.1; JOINED.
 DR EMBL; L32931; AAB59401.1; JOINED.
 DR EMBL; L32932; AAB59401.1; JOINED.
 DR EMBL; L32933; AAB59401.1; JOINED.
 DR EMBL; L32934; AAB59401.1; JOINED.
 DR EMBL; L32935; AAB59401.1; JOINED.
 DR EMBL; L32936; AAB59401.1; JOINED.
 DR EMBL; L32937; AAB59401.1; JOINED.
 DR EMBL; L32938; AAB59401.1; JOINED.
 DR EMBL; L32939; AAB59401.1; JOINED.
 DR EMBL; L32940; AAB59401.1; JOINED.
 DR EMBL; L32941; AAB59401.1; JOINED.
 DR EMBL; L32942; AAB59401.1; JOINED.
 DR EMBL; L32943; AAB59401.1; JOINED.
 DR EMBL; L32944; AAB59401.1; JOINED.
 DR EMBL; L32945; AAB59401.1; JOINED.
 DR EMBL; L32946; AAB59401.1; JOINED.
 DR EMBL; L32947; AAB59401.1; JOINED.
 DR EMBL; L32948; AAB59401.1; JOINED.
 DR EMBL; L32949; AAB59401.1; JOINED.
 DR EMBL; L32950; AAB59401.1; JOINED.
 DR EMBL; L32951; AAB59401.1; JOINED.
 DR EMBL; L32952; AAB59401.1; JOINED.
 DR EMBL; L32953; AAB59401.1; JOINED.
 DR EMBL; L32954; AAB59401.1; JOINED.
 DR EMBL; L32955; AAB59401.1; JOINED.
 DR EMBL; L32956; AAB59401.1; JOINED.
 DR EMBL; L32957; AAB59401.1; JOINED.
 DR EMBL; L32958; AAB59401.1; JOINED.
 DR EMBL; L32959; AAB59401.1; JOINED.
 DR EMBL; L32960; AAB59401.1; JOINED.
 DR EMBL; L32961; AAB59401.1; JOINED.
 DR EMBL; L32962; AAB59401.1; JOINED.
 DR EMBL; L32963; AAB59401.1; JOINED.
 DR EMBL; L32964; AAB59401.1; JOINED.
 DR EMBL; L32965; AAB59401.1; JOINED.
 DR EMBL; L32966; AAB59401.1; JOINED.
 DR EMBL; L32967; AAB59401.1; JOINED.
 DR EMBL; L32968; AAB59401.1; JOINED.
 DR EMBL; L32969; AAB59401.1; JOINED.
 DR EMBL; L32970; AAB59401.1; JOINED.
 DR EMBL; L32971; AAB59401.1; JOINED.
 DR EMBL; L32972; AAB59401.1; JOINED.
 DR EMBL; L32973; AAB59401.1; JOINED.
 DR EMBL; L32974; AAB59401.1; JOINED.
 DR EMBL; L32975; AAB59401.1; JOINED.
 DR EMBL; L32976; AAB59401.1; JOINED.
 DR EMBL; L32977; AAB59401.1; JOINED.
 DR EMBL; L32978; AAB59401.1; JOINED.
 DR EMBL; L32979; AAB59401.1; JOINED.
 DR EMBL; L32980; AAB59401.1; JOINED.
 DR EMBL; L32981; AAB59401.1; JOINED.
 DR EMBL; L32982; AAB59401.1; JOINED.
 DR EMBL; L32983; AAB59401.1; JOINED.
 DR EMBL; L32984; AAB59401.1; JOINED.
 DR EMBL; L32985; AAB59401.1; JOINED.
 DR EMBL; L32986; AAB59401.1; JOINED.
 DR EMBL; L32987; AAB59401.1; JOINED.
 DR EMBL; L32988; AAB59401.1; JOINED.
 DR EMBL; L32989; AAB59401.1; JOINED.
 DR EMBL; L32990; AAB59401.1; JOINED.
 DR EMBL; L32991; AAB59401.1; JOINED.
 DR EMBL; L32992; AAB59401.1; JOINED.
 DR EMBL; L32993; AAB59401.1; JOINED.
 DR EMBL; L32994; AAB59401.1; JOINED.
 DR EMBL; L32995; AAB59401.1; JOINED.
 DR EMBL; L32996; AAB59401.1; JOINED.
 DR EMBL; L32997; AAB59401.1; JOINED.
 DR EMBL; L32998; AAB59401.1; JOINED.
 DR EMBL; L32999; AAB59401.1; JOINED.
 DR EMBL; L33000; AAB59401.1; JOINED.
 DR EMBL; L33001; AAB59401.1; JOINED.
 DR EMBL; L33002; AAB59401.1; JOINED.
 DR EMBL; L33003; AAB59401.1; JOINED.
 DR EMBL; L33004; AAB59401.1; JOINED.
 DR EMBL; L33005; AAB59401.1; JOINED.
 DR EMBL; L33006; AAB59401.1; JOINED.
 DR EMBL; L33007; AAB59401.1; JOINED.
 DR EMBL; L33008; AAB59401.1; JOINED.
 DR EMBL; L33009; AAB59401.1; JOINED.
 DR EMBL; L33010; AAB59401.1; JOINED.
 DR EMBL; L33011; AAB59401.1; JOINED.
 DR EMBL; L33012; AAB59401.1; JOINED.
 DR EMBL; L33013; AAB59401.1; JOINED.
 DR EMBL; L33014; AAB59401.1; JOINED.
 DR EMBL; L33015; AAB59401.1; JOINED.
 DR EMBL; L33016; AAB59401.1; JOINED.
 DR EMBL; L33017; AAB59401.1; JOINED.
 DR EMBL; L33018; AAB59401.1; JOINED.
 DR EMBL; L33019; AAB59401.1; JOINED.
 DR EMBL; L33020; AAB59401.1; JOINED.
 DR EMBL; L33021; AAB59401.1; JOINED.
 DR EMBL; L33022; AAB59401.1; JOINED.
 DR EMBL; L33023; AAB59401.1; JOINED.
 DR EMBL; L33024; AAB59401.1; JOINED.
 DR EMBL; L33025; AAB59401.1; JOINED.
 DR EMBL; L33026; AAB59401.1; JOINED.
 DR EMBL; L33027; AAB59401.1; JOINED.
 DR EMBL; L33028; AAB59401.1; JOINED.
 DR EMBL; L33029; AAB59401.1; JOINED.
 DR EMBL; L33030; AAB59401.1; JOINED.
 DR EMBL; L33031; AAB59401.1; JOINED.
 DR EMBL; L33032; AAB59401.1; JOINED.
 DR EMBL; L33033; AAB59401.1; JOINED.
 DR EMBL; L33034; AAB59401.1; JOINED.
 DR EMBL; L33035; AAB59401.1; JOINED.
 DR EMBL; L33036; AAB59401.1; JOINED.
 DR EMBL; L33037; AAB59401.1; JOINED.
 DR EMBL; L33038; AAB59401.1; JOINED.
 DR EMBL; L33039; AAB59401.1; JOINED.
 DR EMBL; L33040; AAB59401.1; JOINED.
 DR EMBL; L33041; AAB59401.1; JOINED.
 DR EMBL; L33042; AAB59401.1; JOINED.
 DR EMBL; L33043; AAB59401.1; JOINED.
 DR EMBL; L33044; AAB59401.1; JOINED.
 DR EMBL; L33045; AAB59401.1; JOINED.
 DR EMBL; L33046; AAB59401.1; JOINED.
 DR EMBL; L33047; AAB59401.1; JOINED.
 DR EMBL; L33048; AAB59401.1; JOINED.
 DR EMBL; L33049; AAB59401.1; JOINED.
 DR EMBL; L33050; AAB59401.1; JOINED.
 DR EMBL; L33051; AAB59401.1; JOINED.
 DR EMBL; L33052; AAB59401.1; JOINED.
 DR EMBL; L33053; AAB59401.1; JOINED.
 DR EMBL; L33054; AAB59401.1; JOINED.
 DR EMBL; L33055; AAB59401.1; JOINED.
 DR EMBL; L33056; AAB59401.1; JOINED.
 DR EMBL; L33057; AAB59401.1; JOINED.
 DR EMBL; L33058; AAB59401.1; JOINED.
 DR EMBL; L33059; AAB59401.1; JOINED.
 DR EMBL; L33060; AAB59401.1; JOINED.
 DR EMBL; L33061; AAB59401.1; JOINED.
 DR EMBL; L33062; AAB59401.1; JOINED.
 DR EMBL; L33063; AAB59401.1; JOINED.
 DR EMBL; L33064; AAB59401.1; JOINED.
 DR EMBL; L33065; AAB59401.1; JOINED.
 DR EMBL; L33066; AAB59401.1; JOINED.
 DR EMBL; L33067; AAB59401.1; JOINED.
 DR EMBL; L33068; AAB59401.1; JOINED.
 DR EMBL; L33069; AAB59401.1; JOINED.
 DR EMBL; L33070; AAB59401.1; JOINED.
 DR EMBL; L33071; AAB59401.1; JOINED.
 DR EMBL; L33072; AAB59401.1; JOINED.
 DR EMBL; L33073; AAB59401.1; JOINED.
 DR EMBL; L33074; AAB59401.1; JOINED.
 DR EMBL; L33075; AAB59401.1; JOINED.
 DR EMBL; L33076; AAB59401.1; JOINED.
 DR EMBL; L33077; AAB59401.1; JOINED.
 DR EMBL; L33078; AAB59401.1; JOINED.
 DR EMBL; L33079; AAB59401.1; JOINED.
 DR EMBL; L33080; AAB59401.1; JOINED.
 DR EMBL; L33081; AAB59401.1; JOINED.
 DR EMBL; L33082; AAB59401.1; JOINED.
 DR EMBL; L33083; AAB59401.1; JOINED.
 DR EMBL; L33084; AAB59401.1; JOINED.
 DR EMBL; L33085; AAB59401.1; JOINED.
 DR EMBL; L33086; AAB59401.1; JOINED.
 DR EMBL; L33087; AAB59401.1; JOINED.
 DR EMBL; L33088; AAB59401.1; JOINED.
 DR EMBL; L33089; AAB59401.1; JOINED.
 DR EMBL; L33090; AAB59401.1; JOINED.
 DR EMBL; L33091; AAB59401.1; JOINED.
 DR EMBL; L33092; AAB59401.1; JOINED.
 DR EMBL; L33093; AAB59401.1; JOINED.
 DR EMBL; L33094; AAB59401.1; JOINED.
 DR EMBL; L33095; AAB59401.1; JOINED.
 DR EMBL; L33096; AAB59401.1; JOINED.
 DR EMBL; L33097; AAB59401.1; JOINED.
 DR EMBL; L33098; AAB59401.1; JOINED.
 DR EMBL; L33099; AAB59401.1; JOINED.
 DR EMBL; L33100; AAB59401.1; JOINED.
 DR EMBL; L33101; AAB59401.1; JOINED.
 DR EMBL; L33102; AAB59401.1; JOINED.
 DR EMBL; L33103; AAB59401.1; JOINED.
 DR EMBL; L33104; AAB59401.1; JOINED.
 DR EMBL; L33105; AAB59401.1; JOINED.
 DR EMBL; L33106; AAB59401.1; JOINED.
 DR EMBL; L33107; AAB59401.1; JOINED.
 DR EMBL; L33108; AAB59401.1; JOINED.
 DR EMBL; L33109; AAB59401.1; JOINED.
 DR EMBL; L33110; AAB59401.1; JOINED.
 DR EMBL; L33111; AAB59401.1; JOINED.
 DR EMBL; L33112; AAB59401.1; JOINED.
 DR EMBL; L33113; AAB59401.1; JOINED.
 DR EMBL; L33114; AAB59401.1; JOINED.
 DR EMBL; L33115; AAB59401.1; JOINED.
 DR EMBL; L33116; AAB59401.1; JOINED.
 DR EMBL; L33117; AAB59401.1; JOINED.
 DR EMBL; L33118; AAB59401.1; JOINED.
 DR EMBL; L33119; AAB59401.1; JOINED.
 DR EMBL; L33120; AAB59401.1; JOINED.
 DR EMBL; L33121; AAB59401.1; JOINED.
 DR EMBL; L33122; AAB59401.1; JOINED.
 DR EMBL; L33123; AAB59401.1; JOINED.
 DR EMBL; L33124; AAB59401.1; JOINED.
 DR EMBL; L33125; AAB59401.1; JOINED.
 DR EMBL; L33126; AAB59401.1; JOINED.
 DR EMBL; L33127; AAB59401.1; JOINED.
 DR EMBL; L33128; AAB59401.1; JOINED.
 DR EMBL; L33129; AAB59401.1; JOINED.
 DR EMBL; L33130; AAB59401.1; JOINED.
 DR EMBL; L33131; AAB

FT	REPEAT	1302	1310	2-14.
Query Match		8.4%	Score 317;	DB 1; Length 2224;
Best Local Similarity		42.0%	Pred. No. 4.4e-14;	
Matches 71; Conservative		30; Mismatches 52;	Indels 16; Gaps 5;	
QY	253	NGCSKSLSEFE----	PDGQIRASS----	SMQSVNESDQVHMSGQARLDOGSPMSAGDSS 305
DB	2054	NGCSIPPLMEHGKIKINIKOTITASSFFKSSW----	GD--YWEFFRAKLNAQGVNMMOAKA 2116	
QY	306	NNHREWMLEIDLGKRRKTIKRTGTSQSNENFYKSEFVKNKNNKSWKRYKGIYVNE 365		
DB	2117	NNNK--QWLEIDLKIKKITTALITITQGCSSLSSEMYKSTYTHISQGVEMKRYRLKSSMY 2174		
QY	366	EKVFGNSNFRDPVQNNFIPVAVRYRVYVPTWQORLAKVELIGCOI 414		
DB	2175	DKIEFGNNTKGVHNFNFNPITISRFIRYIPKTMQOSITLRLRELFCDI 2223		
RESULT 11				
FA5_PIG	ID	FA5_PIG	STANDARD;	PRT; 2258 AA.
AC	Q9GIP1;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Coagulation factor V precursor (Activated protein C cofactor).			
GN	F5.			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RP	SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C			
RP	DOMAINS.			
RC	TISSUE=Liver;			
RX	MEDLINE=21121490; PubMed=11292814;			
RA	Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,			
RA	Kim H.K.W.;			
RT	"Porcine factor V: cDNA cloning, gene mapping, three-dimensional			
RT	protein modelling of membrane binding sites and comparative anatomy of			
RT	domains.";			
RL	Cell. Mol. Life Sci. 58:148-159(2001).			
CC	-1- FUNCTION: Coagulation factor V is a cofactor that participates			
CC	with factor Xa to activate prothrombin to thrombin.			
CC	-1- SUBUNIT: Factor Va is composed of a heavy chain and a light			
CC	chain, noncovalently bound. The interaction between the two chains			
CC	is calcium-dependent.			
CC	-1- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1			
CC	and C2 may be involved in membrane binding.			
CC	-1- PWM: Thrombin activates factor V proteolytically to the active			
CC	cofactor, factor Va (formation of a heavy chain at the N-			
CC	terminus and a light chain at the C-terminus).			
CC	-1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS, EACH IS COMPOSED OF			
CC	2 PLASTOCYANIN-LIKE REPEATS.			
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.			
CC	-1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF191308; AAC28381.1; -			
DR	HSSP; P12259; ICZT.			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR000421; FA58_C.			
DR	Pfam; PF00394; Cu-oxidase; 3.			
DR	Pfam; PF00754; F5_F8_type_C; 2.			
DR	SMART; SM00231; FA58C; 2.			

DR	PROSITE; PS00079;	MULTICOPPER_OXIDASE1; 2.
DR	PROSITE; PS01285;	FA58C_1; 2.
DR	PROSITE; PS01286;	FA58C_2; 2.
KW	Blood coagulation;	Glycoprotein; Sulfation; Calcium; Signal; zymogen.
KV	Repeat.	
FT	SIGNAL	1 22
FT	CHAIN	23 2258
FT	CHAIN	23 737
FT	PEPTIDE	738 1611
FT	CHAIN	1612 2258
FT	DOMAIN	30 329
FT	DOMAIN	30 193
FT	DOMAIN	203 329
FT	DOMAIN	348 683
FT	DOMAIN	348 525
FT	DOMAIN	535 683
FT	DOMAIN	691 1611
FT	DOMAIN	1168 1539
FT	REPEAT	1168 1176
FT	REPEAT	1177 1185
FT	REPEAT	1186 1194
FT	REPEAT	1195 1203
FT	REPEAT	1204 1212
FT	REPEAT	1213 1221
FT	REPEAT	1222 1230
FT	REPEAT	1231 1239
FT	REPEAT	1240 1248
FT	REPEAT	1249 1257
FT	REPEAT	1258 1266
FT	REPEAT	1267 1275
FT	REPEAT	1276 1284
FT	REPEAT	1285 1293
FT	REPEAT	1294 1302
FT	REPEAT	1303 1311
FT	REPEAT	1312 1320
FT	REPEAT	1321 1329
FT	REPEAT	1330 1338
FT	REPEAT	1339 1347
FT	REPEAT	1348 1356
FT	REPEAT	1357 1365
FT	REPEAT	1366 1374
FT	REPEAT	1375 1383
FT	REPEAT	1384 1392
FT	REPEAT	1393 1401
FT	REPEAT	1402 1410
FT	REPEAT	1411 1419
FT	REPEAT	1420 1428
FT	REPEAT	1429 1437
FT	REPEAT	1438 1446
FT	REPEAT	1447 1455
FT	REPEAT	1456 1464
FT	REPEAT	1465 1473
FT	REPEAT	1474 1482
FT	REPEAT	1483 1491
FT	REPEAT	1492 1500
FT	REPEAT	1501 1509
FT	REPEAT	1510 1518
FT	REPEAT	1519 1527
FT	REPEAT	1531 1539
FT	DOMAIN	1616 1941
FT	DOMAIN	1616 1785
FT	DOMAIN	1795 1941
FT	DOMAIN	1942 2095
FT	DOMAIN	2100 2255
FT	SITE	737 738
FT	SITE	1029 1030
FT	SITE	1611 1612
FT	DISULFID	167 193

FT	DISULFID	499	525	PROBABLE.
FT	DISULFID	1759	1785	PROBABLE.
FT	DISULFID	1941	2095	BY SIMILARITY.
FT	DISULFID	2100	2255	BY SIMILARITY.
FT	MOD_RES	692	692	SULFATION (POTENTIAL).
FT	MOD_RES	696	696	SULFATION (POTENTIAL).
FT	MOD_RES	724	724	SULFATION (POTENTIAL).
FT	MOD_RES	726	726	SULFATION (POTENTIAL).
FT	MOD_RES	745	745	SULFATION (POTENTIAL).
FT	MOD_RES	1560	1560	SULFATION (POTENTIAL).
FT	MOD_RES	1576	1576	SULFATION (POTENTIAL).
FT	MOD_RES	1581	1581	SULFATION (POTENTIAL).
FT	MOD_RES	1584	1584	SULFATION (POTENTIAL).
FT	MOD_RES	1588	1588	SULFATION (POTENTIAL).
FT	MOD_RES	1631	1631	SULFATION (POTENTIAL).
FT	CARBOHYD	153	153	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	225	225	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	467	467	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	741	741	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	752	752	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	760	760	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	776	776	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	782	782	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	899	899	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	960	960	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1048	1048	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1066	1066	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1174	1174	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1480	1480	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1537	1537	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1597	1597	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1737	1737	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1886	1886	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2243	2243	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	2258	2258	N-LINKED (GLCNAC. . .)
FT	SEQUENCE	2258	2258	AA: 256078 MW: 915998E0076A2ACC CMC64:

Query Match 8.3%; Score 314; DB 1; Length 2258;
 Best Local Similarity 39.2%; Pred. No. 7.3e-14;
 Matches 67; Conservative 35; Mismatches 49; Indels 20; Gaps 6;

QY	253	NGCSRLSFE----	PDGQIRASS----	SWQSVNESGDQVHSPQOARLDQG--PSWASGD	303
DB	2098	NGCSTPLGMSGNIKNQIRASSFKKSMW-----	GD--YVEPRARLNAQGRVAMQAKA	2150	
QY	304	SSNNHPRLELDLGGKRTIGRTGSGQSNFNVKSFVANKNNKATYKIVN	363		
DB	2151	NNNN-----OWLDLIDLKIKRTITGCGKSLSEMYKRTTQYSDRGVEMSYRKSS	2206		
QY	364	NEKRVGNSNFRDPYONNFIPIVAVYVAVPQTHORIALKVELIGCOI	414		
DB	2207	MVDKTEGNNNINIGVKNFNPPIISFRITIRKMNQSLALNLELFGCDI	2257		

RESULT 12
 MFGM_HUMAN
 ID MFGM_HUMAN STANDARD; PRT; 387 AA.
 AC 008431:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-BGF factor 8) (MFG-E8) (HMGF)
 DE (Breast epithelial antigen BA46) (MFGM) (Contains: Medin).
 GN MFGEB.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. Breast carcinoma;
 RC TISSUE-Breast, and Breast carcinoma;
 RX MEDLINE=96213908; PubMed=8639264;
 RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
 RT "Cloning and sequence analysis of human breast epithelial antigen
 RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
 RT growth factor-like domain."
 RL DNA Cell Biol. 15:281-286(1996).
 RN [2]
 RP SEQUENCE OF 170-387 FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE=91371351; PubMed=1909932;
 RA Larocca D., Peterson J.A., Drea R., Kunyoshi J., Bistrain A.M.,
 RA Ceriani R.L.;
 RT "A Mr 46,000 human milk fat globule protein that is highly expressed
 RT in human breast tumors contains factor VIII-like domains."
 RL Cancer Res. 51:4994-4998(1991).
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC TISSUE-Milk;
 RX MEDLINE=98194924; PubMed=9535276;
 RA Gufrida M.G., Cavalletto M., Giunta C., Conti A.,
 RA Godovac-Zimmermann J.;
 RT "Isolation and characterization of full and truncated forms of human
 RT breast carcinoma protein BA46 from human milk fat globule membranes."
 RL J. Protein Chem. 17:143-148(1998).
 RN [4]
 RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
 RX MEDLINE=99342076; PubMed=10411933;
 RA Haeggqvist B., Naeslund J., Sletten K., Westermarck G.T., Mucchiano G.,
 RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermarck P.;
 RT "Medin: an integral fragment of aortic smooth muscle cell-produced
 RT lactadherin forms the most common human amyloid."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=97405885; PubMed=9260929;
 RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
 RT Lactadherin (formerly BA46), a membrane-associated glycoprotein
 RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
 RT (RGD)-dependent cell adhesion."
 RL DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to jlicense@sib-sib.ch).
 CC EMBL: U88516; AAC50549.1; -;
 CC EMBL: S56151; AAB19771.1; -;
 CC HSSP: P08709; IBF9.
 CC GeneW: HGNC:7036; MFGEB.
 CC MIM: 602281; -;
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000421; FAS5_C.
 CC Pfam: PF00008; EGF; 1.

RESULT 14
 ID FA8_HUMAN STANDARD; PRT; 2351 AA.
 AC P00451;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor VIII precursor (Procoagulant component)
 DE (Antihemophilic factor) (AHF).
 GN F8 OR F8C.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86081164; PubMed=3935400;
 RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
 RA Hartog K., Kuo C.H., Maslarz F.R., Merryweather J.P., Najarian R.,
 RA Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
 RA Urdas M.S., Valenzuela P., Dahl H.-H.M., Favaloro J., Hansen J.,
 RA Nordfang O., Ezban M.,
 RT "Characterization of the polypeptide composition of human factor
 RT VIII:C and the nucleotide sequence and expression of the human kidney
 RT cDNA.";
 RL DNA 4:333-349(1985).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85061548; PubMed=6438526;
 RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschler J.,
 RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion R.L.,
 RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
 RT "Expression of active human factor VIII from recombinant DNA clones.";
 RL Nature 312:330-337(1984).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85061550; PubMed=6438528;
 RA Tootle J.J., Knopf J.L., Wozney J.M., Soltzman L.A., Buecker J.L.,
 RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
 RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Pass D.N.,
 RA Hewick R.M.;
 RT "Molecular cloning of a cDNA encoding human antihemophilic factor.";
 RL Nature 312:342-347(1984).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93265012; PubMed=1303178;
 RA Gitschler J., Wood W.I.;
 RT "Sequence of the exon-containing regions of the human factor VIII
 RT gene.";
 RL Hum. Mol. Genet. 1:199-200(1992).
 [5]
 RP SEQUENCE OF 2064-2070 FROM N.A.
 RA de Water N.S., Williams R., Browett P.J.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SOLFATION OF TYR-1699.
 RX MEDLINE=91093266; PubMed=1898735;
 RA Leyte A., van Schijndel H.B., Niehs C., Hutner W.B., Verbeet M.P.,
 RA Merens K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor.";
 RL J. Biol. Chem. 266:740-746(1991).
 [7]
 RP SOLFATION.
 RX MEDLINE=92207952; PubMed=1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII.";
 RL Biochemistry 31:3315-3325(1992).
 [8]

RP STRUCTURE BY NMR OF 2322-2343.
 RX MEDLINE=95200924; PubMed=7893714;
 RA Gilbert G.E., Balaja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL Biochemistry 34:3022-3031(1995).
 [9]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE=91221499; PubMed=1902642;
 RA Gitschler J.;
 RT "The molecular basis of hemophilia A.";
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE=89088506; PubMed=2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL Blood 73:1-12(1989).
 [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE=95245332; PubMed=7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 [12]
 RP VARIANT GLN-2326.
 RX MEDLINE=86235434; PubMed=3012775;
 RA Gitschler J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophilic.";
 RL Science 232:1415-1416(1986).
 [13]
 RP VARIANT PRO-2135.
 RX MEDLINE=88096539; PubMed=3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschler J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 [14]
 RP VARIANT GLN-2228.
 RX MEDLINE=88191889; PubMed=2833855;
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 [15]
 RP VARIANT GLY-291.
 RX MEDLINE=88220354; PubMed=2835904;
 RA Youssoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 [16]
 RP VARIANT CYS-1708.
 RX MEDLINE=89274393; PubMed=2499363;
 RA O'Brien D.P., Tuddenham E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL Blood 73:2117-2122(1989).
 [17]
 RP VARIANT CYS-391.
 RX MEDLINE=90001543; PubMed=2506948;
 RA Shima M., Ware J., Yoshioke A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 [18]
 RP VARIANT LEU-189.
 RX MEDLINE=90057680; PubMed=2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;

RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A.";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT LEU-2326.
 RX MEDLINE-89197216; PubMed-2495245;
 RA Inaba H., Fujimaki M., Kazazian H. H. Jr., Antonarakis S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene.";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HIS-391.
 RX MEDLINE-89264602; PubMed-2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H. H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372-->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT CYS-1708.
 RX MEDLINE-90105723; PubMed-2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H. H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANTS GLN-2228 AND LEU-2326.
 RX MEDLINE-90123183; PubMed-2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carobona A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT CYS-391.
 RX MEDLINE-90329422; PubMed-1973901;
 RA Pethlison J.K., Mcvey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site.";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS PHE-1699 AND CYS-1708.
 RX MEDLINE-90152691; PubMed-2105906;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H. H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE-90169988; PubMed-2106480;
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H. H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene.";
 Query Match 7.6%; Score 285.5; DB 1; Length 2351;
 Best Local Similarity 37.6%; Pred. No. 8.2e-12;
 Matches 62; Conservative 32; Mismatches 56; Indels 13; Gaps 4;

DB 2302 FQGNDSFTPVVNSIDPLLTTRYLAIHPOSVWYQIALRMEVLGCE 2346
 RESULT 15
 MFGM_MOUSE
 ID MFGM_MOUSE STANDARD; PRT; 463 AA.
 AC P21956; P97800;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
 DE (Sperm surface protein SP47) (MP47).
 GN MFG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
 RC TISSUE-Mammary gland;
 RX MEDLINE-91046008; PubMed-2122462;
 RA Stubbs J.D., Lekutis C., Singer K.L., Bul A., Yuzuki D.,
 RA Srinivasan U., Parry G.;
 RT "cDNA cloning of a mouse mammary epithelial cell surface protein
 RT reveals the existence of epidermal growth factor-like domains linked
 RT to factor VIII-like sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
 RN [2]
 RP SEQUENCE OF 23-463 FROM N.A.
 RC TISSUE-Testis;
 RA Ensalin M.A.;
 RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
 CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M38337; AAA39534.1; -;
 DR EMBL; Y11684; CAA72380.1; -;
 DR PIR; A36479; A36479.
 DR HSSP; P00740; 1EDM.
 DR MGD; MGI:102768; Mfeg8.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001438; EGF_11.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR PRINTS; PRO0010; EGRBLOOD.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
 FT SIGNAL 1 22
 FT CHAIN 23 463 LACTADHERIN.
 FT DOMAIN 24 61 EGF-LIKE 1.
 FT DOMAIN 64 108 EGF-LIKE 2.
 FT DOMAIN 148 303 F5/8 TYPE C 1.

```

FT DOMAIN 308 463 F5/8 TYPE C 2.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 28 39 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 148 303 BY SIMILARITY.
FT DISULFID 290 294 BY SIMILARITY.
FT CARBOHYD 308 463 BY SIMILARITY.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 426 426 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 30 30 S -> F (IN REF. 2).
FT CONFLICT 35 35 N -> D (IN REF. 1; AA SEQUENCE).
FT CONFLICT 110 147 ETRYNDGEYMTTAVPTAVPTAPPTDLSNNLSR ->
G (IN REF. 2).
Y -> S (IN REF. 2).
H -> T (IN REF. 2).
L -> S (IN REF. 2).
E -> A (IN REF. 2).
SQ SEQUENCE 463 AA; 51465 MW; D78B6CEFBBA724D CRC64;

```

Query Match 7.38; Score 276.5; DB 1; Length 463;
 Best Local Similarity 23.4%; Pred. No. 3.4e-12;
 Matches 112; Conservative 59; Mismatches 118; Indels 189; Gaps 21;

```

QY 94 CASDYLLFTSSDOYGYGSGMTVPKELLANTSEVYVRFESGSHISR-----GF 143
DB 16 CASG--LEAASGD-----FCDS-----SLCLN-----GGTCLTGQDNDIYCLCEPGF 55
QY 144 LITYAS-----SDHPLITCLERASHYLKTEYSKFCPCACRDVAGDISGNM- 189
DB 56 TGLVCHETETGRCSPNCPYDARCLVT-IDTQNGDITETICQCPVYSGIHCETETNY 114
QY 190 -VDG-YRDTSLCKAAL-----HAGIAD----- 211
DB 115 NLDGEYMTTAVPTAVPTAPPTDLSNNLSRCSQTLGMEGAIADSOISASYVMGFM 174
QY 212 ---ELGGQISVLRKGI-----SRYGIIANGVLSRDG----- 241
DB 175 GLQRMGPDELRLRYRTGIVNNAHNASNYDSKPIQVNLKRMVSGVMTQGA-SRAGRAEYL 233
QY 242 -----SLSDKRRLFTSN-----GCSR-- 257
DB 234 KTFKVAISLDGRKFEFIQDESQDKEFLGNDLNNSLKVMNPNLEAOYIRLTPVSGCHRG 293
QY 258 -SLSTFE-----PDGQIRASSSMQSVNESGDQVHWSPGQARLDQDQ 296
DB 294 CTLRFEELGCELCLEPLGLKNNITIPDSQMSASSSKYTNLRA--FGWYPHLGRLDNQG 351
QY 297 P--SNASGDSNNHKKPREWEIDLGKKRTGRTGSGTOSNPNFYKSFVNMFKNNRSK 354
DB 352 KINAM-----TAQSNNAKEMLOVDLGTQROVTLITQGARDEGHIIQYVESYKVAHSDGVQ 407
QY 355 WKTYKGIIVNNEEFVQGNNSNFRDPVONNFIPIVAVRVVVPOTWHORIALKVELIGC 412
DB 408 WTYVE--EGSSSKVFGQNDLNNSHKKNIFKPPMARVAVRLPVSWHNRITLRLELLGC 463

```

Search completed: May 15, 2003, 13:16:06
 Job time : 32.3719 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 69.8563 Seconds
(without alignments)
2108.953 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MYPGARGGALARAARGRL.....YSAPRDLPLNQTAMTALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_ARCHAEA:*
- 2: SP_BACTERIA:*
- 3: SP_FUNGI:*
- 4: SP_HUMAN:*
- 5: SP_INVERTEBRATE:*
- 6: SP_MAMMAL:*
- 7: SP_MHC:*
- 8: SP_ORGANELLE:*
- 9: SP_PHAGE:*
- 10: SP_PLANT:*
- 11: SP_RODENT:*
- 12: SP_VIRUS:*
- 13: SP_VERTEBRATE:*
- 14: SP_UNCLASSIFIED:*
- 15: SP_VIRUS:*
- 16: SP_BACTERIAP:*
- 17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1925.5	51.1	503	11 09D4J3	09d4j3 mus musculu
2	1769.5	46.9	460	11 09D696	09d696 mus musculu
3	1535	40.7	432	11 08R327	08r327 mus musculu
4	1062	28.2	769	11 091ZV3	091zv3 mus musculu
5	1055	28.0	769	11 091ZV2	091zv2 rattus norv
6	1042	27.6	775	4 096PD2	096pd2 homo sapien
7	1042	27.6	775	4 096DX2	096dx2 homo sapien
8	841	22.3	155	4 096NH2	096nh2 homo sapien
9	572.5	15.2	364	4 014089	014089 homo sapien
10	357.5	9.3	926	11 08QZ17	08qz17 mus musculu
11	351.5	9.3	901	4 09H2E4	09h2e4 homo sapien
12	351.5	9.3	901	4 09H2D5	09h2d5 homo sapien
13	351.5	9.3	906	4 09H2E3	09h2e3 homo sapien
14	351.5	9.3	906	4 09H2D4	09h2d4 homo sapien
15	351	9.3	919	13 08UVR0	08uvr0 gallus galli
16	351	9.3	936	13 08UV09	08uv09 gallus galli

17	349	9.3	555	4 09H2E2	09h2e2 homo sapien
18	335	8.9	609	4 096190	096190 homo sapien
19	335	8.9	644	4 0961H5	0961h5 homo sapien
20	335	8.9	704	4 09H2E1	09h2e1 homo sapien
21	325.5	8.6	921	11 09QX38	09qx38 rattus norv
22	321	8.5	2224	4 043737	043737 homo sapien
23	320	8.5	923	13 08QFX6	08qfx6 brachydanto
24	318	8.4	2183	11 088783	088783 mus musculu
25	306.5	8.1	2343	6 018806	018806 canis famli
26	305.5	8.1	2634	6 062730	062730 canis famli
27	299.5	7.9	779	4 09NTT3	09ntt3 homo sapien
28	290	7.7	426	11 09WTS3	09wts3 mus musculu
29	285.5	7.6	216	4 014286	014286 homo sapien
30	279.5	7.4	463	11 09RLX9	09rlx9 mus musculu
31	277	7.3	2119	13 090X47	090x47 brachydanto
32	268	7.1	480	4 043854	043854 homo sapien
33	264	7.0	480	11 035474	035474 mus musculu
34	238.5	6.3	363	6 077718	077718 equus cabal
35	226	6.0	1128	11 088442	088442 mus musculu
36	221.5	5.9	224	11 09RLM6	09rlm6 mus musculu
37	217	5.8	1158	4 014113	014113 homo sapien
38	216.5	5.7	719	11 061281	061281 mus musculu
39	209.5	5.6	728	6 097567	097567 bos taurus
40	200	5.3	1311	4 08WX98	08wx98 homo sapien
41	199.5	5.3	858	5 076470	076470 lytechinus
42	199	5.3	721	4 096M80	096m80 homo sapien
43	194.5	5.2	335	4 09BTL9	09btl9 homo sapien
44	188.5	5.0	745	4 096M75	096m75 homo sapien
45	188.5	5.0	3843	5 09U5D0	09u5d0 drosophila

ALIGNMENTS

RESULT 1	ID	PRELIMINARY	PRT	503 AA.
09D4J3	09D4J3			
AC	09D4J3			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	4631413K1IRIK protein.			
GN	4631413K1IRIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RX	MEDLINE=2108560; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Glass C., King B., Koehle H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schirni L. M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsi G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,			
RA	Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D. A., Kamuya M., Lee N. H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momhaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyo-Oka K., Wang K. H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hashizaki Y. ;			
RT	*Functional annotation of a full-length mouse cDNA collection.*;			
RL	Nature 409:685-690(2001).			
CC	-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.			
DR	EMBL; AK016485; BAB30265.1; -.			

MD: MGI:1913936; 4631413K1R1K.
 DR InterPro: IPR000859; COB_domain.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; COB; 1.
 DR SMART: SM00042; COB; 1.
 DR PROSITE: PS01180; COB; 1.
 SO SEQUENCE 503 AA; 54547 MW; FEB121E845CAA06B CRC64;

Query Match 51.18; Score 1925.5; DB 11; Length 503;
 Best Local Similarity 55.88; Pred. No. 6.9e-140;
 Matches 392; Conservative 34; Mismatches 73; Indels 203; Gaps 4;

14 AAGGGLALLALLANAPRLQAELEGDGGHLYVYDSDSTMTSKNPGTYPHNTEVEKIT 73
 5 AGGVSVALLEAVCAPRLQAELEGDGGHLYVYDSDSTMTSKNPGTYPHNTEVEKIT 64
 74 VPKGRLLRLRGDLDIESQTCADPDLFTSSDDQPGYCGSMWVPEKLLMTSEVTFEE 133
 65 VPKGRLLRLRGDLDIESQTCADPDLFTSSDDQPGYCGSMWVPEKLLMTSEVTFEE 124
 134 SGSHISGKFLITYASSDHPDLITCLERASHYLTETSKFCPCAGCRDVAAGDISGMVDGY 193
 125 SGSHISGKFLITYASSDHPDLITCLERASHYLTETSKFCPCAGCRDVAAGDISGMVDGY 184
 194 RDTSLICAAIHAGITDELGGQISYVORKGISREGILANGVLSRGSLSDKRELFTSN 253
 185 RDTSLICAAIHAGITDELGGHINLQSKISHTYEGILANGVLSRHSLSSEKRLP 241
 254 GCSRLSEFPDQIRASSSMQSVESGDQVHMSPOQARIADQGSWASGSDSNHKKPREW 313
 242 ----- 241
 314 LEIDGEEKKITGTGRTTSGTOSNENFYVSKSVNMFKNNSKWKTKYKGIYNNKEVFGNS 373
 242 ----- 241
 374 NFRDPVQNNFPIVARYRVVPTWHQRIALKEVLLGQITQGDLSLWRTKSSTSVS 433
 242 ----- 241
 434 TKKEDETIRPSEBETSTGINTTVAIPLVLLVLFVAGMIFPAFFKKKKKSPYSA 493
 242 -----TPGNNITTVAILPSYFIALLLTGIGFALCKRRKKKKGPYSA 285
 494 EAQTDCKQIKYPPARHQSAEFTSYDNKEKMTOKDLITSDMADYQOPLMIGTVTR 553
 286 DAQTCCKQIKYPPARHQSTEFITSYDNKEKMTOKDLITSDMADYQOPLMIGTVAR 345
 554 KGSFPRMDTAEAGVSTDAAGHYDCQORAGRHEVALPLAPREVEYTPVERHVALAH 613
 346 KGSFPRMDTAEAGVSTDAAGHYDCQORAGRHEVALPLAPREVEYTPVERHVALAH 405
 614 TFSAGSGRVVPOPGHNSLSGGSFSPVAGVAGDQORPHSAQPADRGYDPRKAVSA 673
 406 TFSQSGRVVPOPGHNSLSGGSFSPVAGVAGDQORPHSAQPADRGYDPRKAVSA 464
 674 LATESGHPDSKPPYHCTSDSYAPRDCLPLAQTATALL 715
 465 L-DSRDPASOSQMT-SGDDGYAPRNGIAPLQNTATALL 503

RESULT 2

Q9D696 PRELIMINARY; PRT; 460 AA.
 AC Q9D696;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE 4631413K1R1K.
 GN 4631413K1R1K protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ichii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Flischnmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki K., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,
 RA Guenincel S., Hill D., Hofmann M., Hume D.A., Kimura M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- STIMULANT: CONTAINS 1 COB DOMAIN.
 DR EMBL: AK014521; BAB29409.1;
 DR MGI: 1913936; 4631413K1R1K.
 DR InterPro: IPR000859; COB_domain.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; COB; 1.
 DR SMART: SM00042; COB; 1.
 DR PROSITE: PS01180; COB; 1.
 SO SEQUENCE 460 AA; 50334 MW; 3CF3356F70A8B93 CRC64;

Query Match 46.98; Score 1769.5; DB 11; Length 460;
 Best Local Similarity 54.48; Pred. No. 6.4e-128;
 Matches 361; Conservative 32; Mismatches 67; Indels 203; Gaps 4;

53 MTSKNYPTGYPHNTEVEKITVPPKGRLLRLRGDLDIESQTCADPDLFTSSDDQPGY 112
 1 MTSKNYPTGYPHNTEVEKITVPPKGRLLRLRGDLDIESQTCADPDLFTSSDDQPGY 60
 113 GSWTVPEKLLMTSEVTFEESSGSHISGKFLITYASSDHPDLITCLERASHYLTETSK 172
 61 GSWTVPEKLLMTSEVTFEESSGSHISGKFLITYASSDHPDLITCLERASHYLTETSK 120
 173 FCPAGCRDVAAGDISGMVDGYRDTSLICAAIHAGITDELGGQISYVORKGISREGIL 232
 121 FCPAGCRDVAAGDISGMVDGYRDTSLICAAIHAGITDELGGHINLQSKISHTYEGIL 180
 233 ANGVLSRHSLSSEKRLP 198
 181 ANGVLSRHSLSSEKRLP 198
 293 QDQGSWASGSDSNHKKPREWLEIDGEEKKITGTGRTTSGTOSNENFYVSKSVNMFKNNS 352
 199 ----- 198
 353 SKWKTKYKGIYNNKEVFGNSNFRDPVQNNFPIVARYRVVPTWHQRIALKEVLLG 412
 199 ----- 198
 413 QITQGDLSLWRTKSSTSVTKKEDETIRPSEBETSTGINTTVAIPLVLLVLFVA 472
 199 -----TPGNNITTVAILPSYFIALLLT 221
 473 GMSGIFAFRRKKKKKSPYSAEAOCTDCKQIKYPPARHQSAEFTSYDNKEKMTOKDL 532
 222 GMSGIFAFRRKKKKKSNPVSAOCTDCKQIKYPPARHQSTEFITSYDNKEKMTOKDL 281
 533 ITSDMADYQOPLMIGTVTRKGSFPRMDTAEAGVSTDAAGHYDCQORAGRHEVALP 592

Db 282 ITSDMADYQOPLMIGTGTAVARKSGTFRRMDTTEEVNVTASGCHYDCPHRGRHEVALP 341
 QY 593 LAPPEPYATPIVERHYLRAHTFSAOSGYRVPQPGHKHSLSSGFSPPVAGVAGQDGDY 652
 Db 342 LTHSEPEYATPIVERHYLRAHTFSAOSGYRVPQPGHKHSLSSGFSPPVAGVAGQDGDY 400
 QY 653 QPHSAQPADRGYDRPRAVASALATESGHPDSQRPPTHPGTSDSYAPRDCLTPLNOTAMT 712
 Db 401 QRPASPKPVGGYDKPAASSFL--DSRDPASQSQMT--SGDDGYSAAPRNLAPLNOTAMT 457
 QY 713 ALL 715
 Db 458 ALL 460

RESULT 3

Q8R327 PRELIMINARY; PRT; 432 AA.
 ID Q8R327
 AC Q8R327;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Similar to RIKEN CDNA 4631413k11 gene.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL EMBL, BC026771; AAH26771.1;
 SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 40.7%; Score 1535; DB 11; Length 432;
 Best Local Similarity 47.0%; Pred. No. 6.9e-110;
 Matches 330; Conservative 26; Mismatches 72; Indels 274; Gaps 5;

QY 14 AAGRGTLALLAVSAPRLQAEELGDCGHLVYQDSGTMTSKNYPGTYPNHTVCERTIT 73
 Db 5 AGPSYVALLFAYCAPRLQAEEL----- 28
 QY 74 VPKGRLLRLGLDIESQTCASDYLLFTSSDQYGYGSMVTPKELLNTSEVTVFE 133
 Db 29 -----GPGCSMAVPKRLRLNSNEVTVLFK 53
 QY 134 SSGHISGRGLLTYASSDHDITCLERASHYLTKEYSKRCPCAGCRVADISGMVGY 193
 Db 54 SSGHISGRGLLTYASSDHDITCLERASHYLTKEYSKRCPCAGCRVADISGMVGY 113
 QY 194 RQTSILCKAIIHAGIITADELGGQISVLRKGISREYGLANGVLSRDSISDKRFLPTSN 253
 Db 114 RQTSILCKAIIHAGIITADELGGQISVLRKGISREYGLANGVLSRDSISDKRFLPTSN 170
 QY 254 GCSRSLSFEPDGOIRASSSQSVNESGQYHWSFGARLDQGPSSMNSNHRKPREW 313
 Db 171 ----- 170
 QY 314 LEIDLGEKKITIGRTGTQSNFNFYKSFVNFKNNSKMTYKGIANNNEKYFQGS 373
 Db 171 ----- 170
 QY 374 NFRDPVQNNFIPPIVARYVAVPQTHORIALKVELIGCQITQGNDSLAVNRKTSQSIVS 433
 Db 171 ----- 170
 QY 434 TKKEDETTRPISEETSTGINTTVAIPVLVVLVFAAGMIFAAFRKKKKKSPGSA 493
 Db 171 -----TPPMNITVAIPSVIFALLTGGIAPICRKRKKGNPYSA 214
 QY 494 EAKTQCKQIKYKPAHOSAEFTISYDNKEMOKDLITSDMADYQOPLMIGTGTAVR 553
 Db 215 DAQKTCWKQIKYKPAHOSAEFTISYDNKEMOKDLITSDMADYQOPLMIGTGTAVR 274

QY 554 KSTFRPMDDTAEAGVSTDAAGHYDCPORAGREHYALPLAPPEYATPIVERHYLRAH 613
 Db 275 KSTFRPMDDTAEAGVSTDAAGHYDCPORAGREHYALPLAPPEYATPIVERHYLRAH 334
 QY 614 TFSAGSYRVPQPGPGHKHSLSSGFSPPVAGVAGQDGDYQPHSAQPADRGYDRPRAVSA 673
 Db 335 TFSAGSYRVPQPGPGHKHSLSSGFSPPVAGVAGQDGDYQPHSAQPADRGYDRPRAVSA 393
 QY 674 LATESGHPDSQRPPTHPGTSDSYAPRDCLTPLNOTAMTALL 715
 Db 394 L--DSRDPASQSQMT--SGDDGYSAAPRNLAPLNOTAMTALL 432

RESULT 4

Q912V3 PRELIMINARY; PRT; 769 AA.
 ID Q912V3
 AC Q912V3;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-ICR;
 RC PubMed=11447234;
 RX Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
 RA Kusunoki A., Sasayama S., Honjo T., Tashiro K.,
 RT "ESDN", A Novel Neuropilin-Like Membrane Protein Cloned from Vascular
 RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
 RT Up-regulated after Vascular Injury."
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL: AE387548; AL30179.1; -.
 DR MGD: MG11920629; ESDN.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Query Match 28.2%; Score 1062; DB 11; Length 769;
 Best Local Similarity 34.9%; Pred. No. 4.9e-73;
 Matches 275; Conservative 121; Mismatches 274; Indels 118; Gaps 23;

QY 3 PGARGGALARAAGRGL-----ALLAVSAPRLQAEELGDCGHL 44
 Db 13 PGARGGALARAAGRGL-----ALLAVSAPRLQAEELGDCGHL 72
 QY 45 VTQDSGTMTSKNYPGTYPNHTVCERTITPKGRLLRLGLDIE--SQTCASDYLTFT 102
 Db 73 VLPESGTLISNYPHTYPSNVCWEIIRVTEGRIRIKFQDFIEDSDCHLNYLKFEN 132
 QY 103 S-----SDQYGYG--SMTVPRELLNTSEVTVRESGSHISGRFLITYASDHPDLITC 158
 Db 133 GIGVSRTEIGYCGLGQNMOSIESKGEVYLFMSSTHAAGRFLASVYIDKEDLITC 192
 QY 159 LERASHYLTKEYSKRCPCAGCRVADISGMVGYRDTSLCKAIIHAGIITADELGGQIS 218
 Db 193 LDTVSNLPEPFSKYKPCAGCLPRAELISGTLPHGYRDSPLCMGIIHAGIIVSNLGGQIS 252
 QY 219 VLQRKGISREYGLANGVLSRDSISDKRFLPTSNNGCSRSLSFE----PDQIRASSSQW 274
 Db 219 VLQRKGISREYGLANGVLSRDSISDKRFLPTSNNGCSRSLSFE----PDQIRASSSQW 274

Db 253 IVISGTPYEYSSLANNTSTVGYLSASLFFPKTSGCYGLMGESGVADPOITASSALE 312
 Qy 275 SVNESCDOVHMSPGQARLODOPSWASGDSNNHKKPREMLTIDLEKKKITGRTGSGTO 334
 Db 313 WTDHNGGNSWAEKARLRKPPPMWA -PATDEH---QWLOIDLNKREKKTITVTTGSGM 368
 Qy 335 SNFNFPYKSPVNNFNKNNKSKWKTYGIVNNEKEVFOGNSNFRDPVONNEIPPIVARYRV 394
 Db 369 IEHSYVAVRYLSDGQRMVYREPGVDQKIFQGNNDYKHDVNNLPPIIARFIV 428
 Qy 395 VQTHORHATLKVLEIGCOIT-----QGNDSLVWRKTS----- 427
 Db 429 NEVQOQKIAMVELLGCQFTLKGRLPKITPPRRNGNN---LRNTARPKLKGRAPKFT 485
 Qy 428 ----QSTSVSTKKEDEITTRPIPESETSTGINTTVAIPVLVAVFAGMGIPAA-- 480
 Db 466 QVLOGRSNNELPVQPAEFTTTPDINKNTVTPSVTKDVALAANLVYLVVALTLLILVLC 545
 Qy 461 -----RKKKKKSPYGSNAOKTDCWKQIKYF-----ARHQAEEFTISYDNE 523
 Db 546 AMHMRNRKKKTEGA-YDLPHMDRACMMKGMKOLLPAKSVDBHEETPVRYSTSE--VSHLSA 602
 Qy 524 KEMTOKLIDITSMADYQOPLMIGTGYTRKSGTFRPMOTDAEBAG-VSTDAGCHYDCQ 582
 Db 603 REVT---TVLODSAEVQAPLVGIVGTIHOIRSTERP--EEGKEAGYADLDP--YNSPM 654
 Qy 583 RAGREHEVALPLAPPEYATPIV---ERHVLRAHTFSAOSGYRVPQPGHKHSLSSGCF 639
 Db 655 QEVYAVYAEPLPVYGTPEYATPIVMDMSGHPASVGLPSTSTFKTAGTQ---HAL----- 706
 Qy 640 SFVAGVGAODDGYQPHSAOPADRGYDRPKAVSALATES---GHPDSQKPTTHPGTSDS 695
 Db 707 ----VGVTYNTLLSTKSDSSGQAOYDTPKGGKSAATPEELVYOVPQSTQELSGARDEK 761
 Qy 696 YSAPRDL 703
 Db 762 FDAKEIL 769

RESULT 5
 Q912V2 PRELIMINARY: PRT: 769 AA.

ID 0912V2
 AC 0912V2
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-1-like protein.
 GN ESDN.
 OS Rattus norvegicus (rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX PubMed=11447234;
 RA Kuboke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the longest Secretory Signal Sequence among Eukaryotes, Is Up-regulated after Vascular Injury."
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1 SIMILARITY: BELONGS TO THE CRP/FMR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 CC EMBL: AF387549; AAL30180.1; -.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.

SQ SEQUENCE 769 AA; 83915 MW; 587936C8BA063A93 CRC64;
 Query Match 28.0%; Score 1055; DB 11; Length 769;
 Best Local Similarity 35.1%; Pred. No.1-7e-72;
 Matches 277; Conservative 114; Mismatches 278; Indels 120; Gaps 25;

Qy 3 PGARGGALARAARG------LALLAVSAPRLQAEEGLD 39
 Db 13 PDDCGRAAPATGRAPLPASAGMCPGPCRNSSRRLLLLLLLLPDA-----GAQKOD 67
 Qy 40 GCGHLVYTDGSGTSMKPNPGYPMHTVCEKITYPKKRLRLRGDDIE-SQRCASDY 98
 Db 68 GCGHTVLPESGTLINPHYPMNVYCKWEIRKTKTGIRIKGDDIEDSDYCHLNY 127
 Qy 99 L-LFSS---SSDQYPCG-SMTVPKELLNTESEVTVREPSGSHSGREFLLTYASDDP 153
 Db 128 LKIFNGIGYSREIKGYCGLGLOMNSIESKSEITVLFMSIHASGRFLASYIDKQ 187
 Qy 154 DLITCLERASHYLTKEYSCPCAGCRDVAAGDISGMNVGDRDTSLCKNAIHAGITADEL 213
 Db 188 DLITCLDVSNFLPEPESKYPACGLLPFAEISGTPHGRDSSPLCMAGIHAGVSDVL 247
 Qy 214 GGOISVLOKKGISREGLIANGVLSRDGSLDKRFLFNSGSRSLSPF---PQOIRA 269
 Db 248 GGQISVYISKGPYEYSSLANNTSMVGLSTSLFTFKTSGCYGLMGESGVADPOQITA 307
 Qy 270 SSSWOSVNESGDOVHMSPGQARLODOPSWASGDSNNHKKPREMLTIDLEKKKITGIRT 329
 Db 308 SSVLETDHMDGENSEMKPKARLRKPPPMWA -PATDEH---QWLOIDLNKREKKTITGVT 363
 Qy 330 TGSSTOSNENFYKSFVNNFNKNNKSKWKTYGIVNNEKEVFOGNSNFRDPVONNEIPPIYA 389
 Db 364 TGSSTLEHNYVAVRYLSDGQRMVYREPGVDQKIFQGNNDYKHDVNNLPPIIARFIV 423
 Qy 390 RYVRVYQTHORHATLKVLEIGCOIT-----QGNDSLVWRKTSQSTSVSTR 435
 Db 424 REIRVNPVOQKIAMVELLGCQFTLKGRLPKITOPPPRRNGNN---KNITVHRKLGRA 481
 Qy 436 KDEDITTRP-----IPSEETST-GINITT-----VAIPVLVAVFAGMGIFA 478
 Db 482 PKFTQALQPSRNDLPLPAQTATPDVKNVTVPSTVKDVALAANLVYLVVALTLLILVLC 541
 Qy 479 AF-----RKKKKKSPYGSNAOKTDCWKQIK--YFAHQSAEFTISYDNE----- 523
 Db 542 ILVCAHMRNRKKKKAEGT-YDLPHMDRACMMKGMKOLLPAKSVDBHEETPVRYSTSEHL 600
 Qy 524 --KEMTOKLIDITSMADYQOPLMIGTGYTRKSGTFRPMOTDAEBAGYSTAGHYDCP 581
 Db 601 SPREVT---TVLODSAEVQAPLVGIVGTIHOIRSTERP--EEGKEASTYADLDP---YANP 653
 Qy 582 QARREHEVALPLAPPEYATPIV---ERHVLRAHTFSAOSGYRVPQPGHKHSLSSGCF 638
 Db 654 VQEVYAVYAEPLPVYGTPEYATPIVMDMSGHPASVGLPSTSTFKTAGTQ----- 702
 Qy 639 SFVAGVGAODDGYQPHSAOPADRGYDRPK-AVSALATES---GHPDSQKPTTHPGTSD 694
 Db 703 --PPALVGTNTLLSTKSDSSGQAOYDTPKGGKSAATPEELVYOVPQSTQELSGARDE 760
 Qy 695 YSAPRDL 703
 Db 761 KFDAKEIL 769

RESULT 6
 Q96PD2 PRELIMINARY: PRT: 775 AA.

ID 096PD2
 AC 096PD2
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-1-like protein.
 GN ESDN.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed-11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
 RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT *ESDN, a Novel Neuropeptide-like Membrane Protein Cloned from Vascular
 RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
 RT Up-regulated after Vascular Injury.*
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL: AF387547; AAL30178.1; -;
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000421; PA58_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR ProSITE: PS01180; CUB; 1.
 KM DNA-binding; Transcription regulation.
 SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Query Match 27.6%; Score 1042; DB 4; Length 775;
 Best local Similarity 36.4%; Pred. No. 1.7e-71;
 Matches 263; Conservative 113; Mismatches 240; Indels 106; Gaps 23;

20 LALLAVSAPLRQAELDGCGLVYDSDGTMKNTPGYRNHTVCERTVPRGR 79
 111 LALLVLLLEBAGAQDGCGLVYDSDGTMKNTPGYRNHTVCERTVPRGR 110
 51 LALLVLLLEBAGAQDGCGLVYDSDGTMKNTPGYRNHTVCERTVPRGR 110

80 LILRLDLDIE-SQTCASDYLFTS-----SSDQYGYCG-SMTVPRELLNTSEVTARE 133
 111 VRIKEDFDIEDSDCHFNRLRYNGIGVSRTEIGYCGIGLGMNHSIESKNEITLFW 170
 134 SSSHSIGRGFLTYASSDHPDLITCLERASHYKTEYSKFCPCAGCDVADISGNVVDY 193
 171 SGHVSGRGFLASYVIDKODLITCLDTASNFLEPEFSKPCAGCLLPFAEISGTIPHG 230
 194 ROTSLLCKAIIHAGITADELGGQISVLRKGISRYEGIIANGVLSHDGSLDRFLFTSN 253
 231 ROSSPCLMGAVHAGVSNLTGGQISVLRKGISRYEGIIANGVLSHDGSLDRFLFTSN 290
 254 GCSRSLSFE---PDQIRASSWSQSVNESGDQVHNSPGQARLQDQPSWASGSSNNHK 309
 291 GCYGTIGMSGVYADPQITASSVLENTDHTGQENSRRPKARLKKRGPWMAAFATDE--- 347
 310 PREMLTIDIGERKKITGITTSQSNFNFYKSFYMANFRKNNNSKKTYKGIYNNEEKYF 369
 348 -YQWLOIDLNKREKKTIGITTSQSNFNFYKSFYMANFRKNNNSKKTYKGIYNNEEKYF 406
 370 OGSNSFRDVPQNNFIPPIYARVYVPQTHORIALKVELIGQ-----ITQ----- 416
 407 OGNKVDHGVNNFPLPIARFIRVNPQMOQKIAMKMLLGGQIFPKGRPRKLTQPPPP 466
 417 --GNSLVRKTSQTSVSTKKEDEITRPI-----PSEET--STGINITT----- 458
 467 RNSND-----LKNITAPPKIAKGRAPKFTQPLPRSSNEFPATQTEOTASPDINTVTPN 522
 459 ----VAIPVLVLLVFAQMGIFAFAF-----RRKKRKS---PYGSAEAKQTKCMQ 503
 523 VTKDVALAALVAVLVMVLTLLILLIVCAMHMRNRKKTGEGYDLPY-----DRAGWMKG 578
 504 IKYFP-----ARQSAEFTISYDNEKEMQKDLISDMADYQOPLMIGGYTA 552
 579 MKQPLPAKAVDHEETPVRSSE--VNHLSPREV--TVLADASEYAOPLGVIGTL 633
 553 RKGSTFRPMDTAEAG-VSTDAGHYDCPORAGREYVALPAPPEEATPV---ERR 608
 634 HGRSTFKP--EGCKEAGYADLP---YNSPGQEVYHAIVAEPLPITQPEYATPILMMSGH 688

QY 609 VLRAHTFSAGSYRVDGPQPGHKNHSLSSGGSPVAGVAGADGDYORPHSAQPADRGYDR 668
 DB 689 PRTSGVGPSTSTFKKANGNP-----PPL--VGTYNTLLSTNDSSSSAQADYDR 735

QY 669 KA 670
 DB 736 KA 737

RESULT 7
 08TDX2
 ID 08TDX2 PRELIMINARY; PRT; 775 AA.
 AC 08TDX2;
 DT 01-JUN-2002 (TREMREL. 21, Created)
 DT 01-JUN-2002 (TREMREL. 21, Last sequence update)
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
 DE C1CPL.
 GN C1CPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
 RA Tatematsu Y., Mitsuuchi T., Nakao A., Takahashi T.;
 RT *Significant up-regulation of a novel gene, C1CPL, in a highly
 RT metastatic lung cancer subline as well as in lung cancers in vivo.*
 RL Oncogene 0:0-0(2002).
 DR EMBL: AB073146; BAB91138.1; -;
 SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Query Match 27.6%; Score 1042; DB 4; Length 775;
 Best local Similarity 36.4%; Pred. No. 1.7e-71;
 Matches 263; Conservative 113; Mismatches 240; Indels 106; Gaps 23;

20 LALLAVSAPLRQAELDGCGLVYDSDGTMKNTPGYRNHTVCERTVPRGR 79
 51 LALLVLLLEBAGAQDGCGLVYDSDGTMKNTPGYRNHTVCERTVPRGR 110

80 LILRLDLDIE-SQTCASDYLFTS-----SSDQYGYCG-SMTVPRELLNTSEVTARE 133
 111 VRIKEDFDIEDSDCHFNRLRYNGIGVSRTEIGYCGIGLGMNHSIESKNEITLFW 170
 134 SSSHSIGRGFLTYASSDHPDLITCLERASHYKTEYSKFCPCAGCDVADISGNVVDY 193
 171 SGHVSGRGFLASYVIDKODLITCLDTASNFLEPEFSKPCAGCLLPFAEISGTIPHG 230
 194 ROTSLLCKAIIHAGITADELGGQISVLRKGISRYEGIIANGVLSHDGSLDRFLFTSN 253
 231 ROSSPCLMGAVHAGVSNLTGGQISVLRKGISRYEGIIANGVLSHDGSLDRFLFTSN 290
 254 GCSRSLSFE---PDQIRASSWSQSVNESGDQVHNSPGQARLQDQPSWASGSSNNHK 309
 291 GCYGTIGMSGVYADPQITASSVLENTDHTGQENSRRPKARLKKRGPWMAAFATDE--- 347
 310 PREMLTIDIGERKKITGITTSQSNFNFYKSFYMANFRKNNNSKKTYKGIYNNEEKYF 369
 348 -YQWLOIDLNKREKKTIGITTSQSNFNFYKSFYMANFRKNNNSKKTYKGIYNNEEKYF 406
 370 OGSNSFRDVPQNNFIPPIYARVYVPQTHORIALKVELIGQ-----ITQ----- 416
 407 OGNKVDHGVNNFPLPIARFIRVNPQMOQKIAMKMLLGGQIFPKGRPRKLTQPPPP 466
 417 --GNSLVRKTSQTSVSTKKEDEITRPI-----PSEET--STGINITT----- 458
 467 RNSND-----LKNITAPPKIAKGRAPKFTQPLPRSSNEFPATQTEOTASPDINTVTPN 522
 459 ----VAIPVLVLLVFAQMGIFAFAF-----RRKKRKS---PYGSAEAKQTKCMQ 503
 523 VTKDVALAALVAVLVMVLTLLILLIVCAMHMRNRKKTGEGYDLPY-----DRAGWMKG 578

```

QY 504 IKYFP-----ARHQAETISYDNKEKMTOKLIDLTSMADYQOPLMIGTGT 552
DB 579 MKOFLPAKAVDHEEPYSSSE--VNHLSPREVT---TVLQADSAEYAGLVGIVGTL 633
QY 553 RKGSPFRMDVDAEAG-VSTDAGGHYDCPQAGHHEVALPLAPPEPYATPIV---ERH 608
DB 634 HORSTFKP--BEKGAGYADIDP--YNSPGOEYHAALPEPTGPAPPIIDMSGH 688
QY 609 VLRAHFSAGSYRVPQPGHSHSLSSGFSYVAGVAGQDGYQRPASHQAPADRGYDRP 668
DB 689 PTVSGVQSPSTFPAKATGNP-----PPL--VGTWTLNLSRMDSCSSAQAYDTF 735
QY 669 KA 670
DB 736 KA 737

RESULT 8
QY 096NH2 PRELIMINARY; PRT; 155 AA.
AC 096NH2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CDNA FLJ30900 f1s, clone FEBR2005752.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibata T.;
RT "unpublished";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; D29810; BA018909.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA5B_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_Type_C; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW Hypothetical protein.
FT NON TER
SO SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Query Match 15.28; Score 572.5; DB 4; Length 364;
Best Local Similarity 38.68; Pred. No. 8.1e-36;
Matches 147; Conservative 61; Mismatches 124; Indels 49; Gaps 16;

QY 39 DCGCHLYWDSGIMTSKNYPGTYPNHVCYCKITTVPRKRLILRLDLDIE-SQTGASD 97
DB 1 DCGCHLYWDSGIMTSKNYPGTYPNHVCYCKITTVPRKRLILRLDLDIE-SQTGASD 97
QY 98 YLFTS-----SSDYGPGYCG-SMTVPKELLNLTSEVTVRFESGSHISORGFLLYASDH 152
DB 61 YLRIYNGVGRTEIGKCYGIGLQINSHIESKNEITLLFMSGHVSGRGLAYSVID- 119
QY 153 PDLITCLERASHYKTEYSKF-----CPACGRVADIGSNMVDGYRDTSLCKAA 203
DB 120 -----KORSNYLFCHGICQIFFMNLSVSTAPACLLPFAISGTFPHGRDSSPLCMAG 172
QY 204 IHAGLIDELGQISYLRKGISRYEGILANGVLSROSL-SDKRLFTSNGCGRSLFE 262
DB 173 VHAGVSTLGLGQISYLRKGISRYEGILANGVLSROSL-SDKRLFTSNGCGRSLFE 262
QY 263 PDGQIRASS-----WQSVNESGDQVHWSGQARLDQGPSW-ASGDSNNKPREWL 314
DB 233 SGCD-RGSNNSTIVLEH--IDHNGQENSMKPKRSQAE--KWTALGAFATDE--YQWL 284
QY 315 EIDGKKKKTIGITGSGTQ-SNENFYYSK--FVNFKNKNSKWKYKGIYNNKVFQ 370
DB 285 QIDNKKKKITGITGSGTQ-SNENFYYSK--FVNFKNKNSKWKYKGIYNNKVFQ 370
QY 371 GNSN-FRDPVONFIPIVAR 390
DB 341 GNRKTRIRVNRNFPPIIAR 361

RESULT 10
QY 08QZV7 PRELIMINARY; PRT; 926 AA.
AC 08QZV7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Neurofilin-2(al7).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-11S, AND ISS.
RC MEDLINE=21363810; PubMed=11471062;
RA Entinger M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related QTLs";
Mamm. Genome 12:657-663(2001).
RN [2]

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Shibata T.;
RT "unpublished";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; D29810; BA018909.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA5B_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_Type_C; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW Hypothetical protein.
FT NON TER
SO SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Query Match 15.28; Score 572.5; DB 4; Length 364;
Best Local Similarity 38.68; Pred. No. 8.1e-36;
Matches 147; Conservative 61; Mismatches 124; Indels 49; Gaps 16;

QY 39 DCGCHLYWDSGIMTSKNYPGTYPNHVCYCKITTVPRKRLILRLDLDIE-SQTGASD 97
DB 1 DCGCHLYWDSGIMTSKNYPGTYPNHVCYCKITTVPRKRLILRLDLDIE-SQTGASD 97
QY 98 YLFTS-----SSDYGPGYCG-SMTVPKELLNLTSEVTVRFESGSHISORGFLLYASDH 152
DB 61 YLRIYNGVGRTEIGKCYGIGLQINSHIESKNEITLLFMSGHVSGRGLAYSVID- 119
QY 153 PDLITCLERASHYKTEYSKF-----CPACGRVADIGSNMVDGYRDTSLCKAA 203
DB 120 -----KORSNYLFCHGICQIFFMNLSVSTAPACLLPFAISGTFPHGRDSSPLCMAG 172
QY 204 IHAGLIDELGQISYLRKGISRYEGILANGVLSROSL-SDKRLFTSNGCGRSLFE 262
DB 173 VHAGVSTLGLGQISYLRKGISRYEGILANGVLSROSL-SDKRLFTSNGCGRSLFE 262
QY 263 PDGQIRASS-----WQSVNESGDQVHWSGQARLDQGPSW-ASGDSNNKPREWL 314
DB 233 SGCD-RGSNNSTIVLEH--IDHNGQENSMKPKRSQAE--KWTALGAFATDE--YQWL 284
QY 315 EIDGKKKKTIGITGSGTQ-SNENFYYSK--FVNFKNKNSKWKYKGIYNNKVFQ 370
DB 285 QIDNKKKKITGITGSGTQ-SNENFYYSK--FVNFKNKNSKWKYKGIYNNKVFQ 370
QY 371 GNSN-FRDPVONFIPIVAR 390
DB 341 GNRKTRIRVNRNFPPIIAR 361

RESULT 10
QY 08QZV7 PRELIMINARY; PRT; 926 AA.
AC 08QZV7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Neurofilin-2(al7).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-11S, AND ISS.
RC MEDLINE=21363810; PubMed=11471062;
RA Entinger M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related QTLs";
Mamm. Genome 12:657-663(2001).
RN [2]

```


QY 537 MADYQPLMIGTGTTRKSTFRPMDTDAEAGVSTDAAG 575
DB 594 WTD-SKPTVETLGPVKSEETTPPTTEE-----ATECG 627

RESULT 12

09H2D5 PRELIMINARY; PRT; 901 AA.

AC 09H2D5; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Neuropilin-2b(0).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20564205; PubMed-11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes;
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF281074; AAG41899.1; -.
DR HSSP; P12259; 1CZT.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00060; MAM_2; 1.
DR GlycoProfile; Neurone; Receptor; Transmembrane.
KW SEQUENCE 901 AA; 101381 MW; A1310304AFD086AB CRC64;

Query Match 9.3%; Score 351.5; DB 4; Length 901;
Best Local Similarity 22.2%; Pred. No. 3.6e-18;
Matches 142; Conservative 99; Mismatches 25; Indels 143; Gaps 25;

QY 41 CGHATYDSDGTMSTSKNPGYVPHNTVEKTIYVK-GKRLRLRG-DLDESQCASD 98
DB 28 CGGRNLSMDAGYITSPGPDYPSHONCEMYIAPBPQKILANPPELEKHKCKYDF 87
QY 99 LIF-----TSSSDQGYPGYSGMTVPKELLANTSEYVRESGSHISGRFLTY-----AS 149
DB 88 IEIRGDSESDLLGKHCNGI-APPTIISGSMYIKFTSDVARGAGFSLRYELFKGS 146
QY 150 SD-----HDLITC-----LERASHYLKTEYSK----- 173
DB 147 EDCSKNFTSPNGTIESPGFPKPYPHNLDCTTILAKPKMEIILQPLIDLEHPLQVGG 206
QY 174 -CPAGCRVDVADIS--GNMVGYRFTSLCKAKAIHAGIYADELGQISVLQRKGSRYEG 230
DB 207 DCKIDMDIWMGICIRAVGLIGYCTKTPSELRSSTGLSLPFRHDMVAK-----DG 259
QY 231 ILANGVLSRDGSLDKRFLFTSNGCSRSLSEF---PDGIRASSSMQSVESGQDVHMS 286
DB 260 FSARVYLIVHQBLENFQ-----CNVPLGMSGRIANEQISASTY-----SDGKWT 305

QY 287 PQQARLQDQGSWASGSSNNHKKPREMLEIDGKKKTTGRTTG--STQSNFNFYKSF 344
DB 306 PQOSRLHGDNDGWTPENLDSN-----KEYLDVDFRLTMTLALATQCAISRETONGYVKS 361
QY 345 VVNFKNNSKMTKYGIYVNEEKVFGNSNFRDPVQNNFIPYVARYVVPOTFHQIA 404
DB 362 KLEVSTNGEDMKVYTRHGKNN--KYQANNDATEVVLNLAHLRFRVIRPQTHSGIA 419
QY 405 LKVELIGQITQGNDSLVRKTS---QSTSVSTKKEDETITRP-----IP 446
DB 420 LRLEIFGRVTDAPCSNMLGMLSLIADQSISASTQGLYMLSPSARLVSSRSGWFRIP 479
QY 447 SEETSGINITTVALPVLVLYVFAAGC---IFA-----ARKKKKSGPYGSAEAQXTD 499
DB 480 QAQPEEMLYVDLGPVKTVGVIIQARGGDSITAVEARAFVRKFKVSYSLNGXD----- 534
QY 500 CWKQIKYPEFARHQA-EETISYD-----NEKEMTKDLINSD 536
DB 535 -WEYIDPRTPQPKLFEENMHYDTPDIRFPDIPAQYVYVPERNSPAGICNRLEVLGCD 593
QY 537 MADYQPLMIGTGTTRKSTFRPMDTDAEAGVSTDAAG 575
DB 594 WTD-SKPTVETLGPVKSEETTPPTTEE-----ATECG 627

RESULT 13

09H2E3 PRELIMINARY; PRT; 906 AA.

AC 09H2E3; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Neuropilin-2b(5).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20564205; PubMed-11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes;
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF280545; AAG41404.1; -.
DR HSSP; P12259; 1CZT.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00060; MAM_2; 1.
DR GlycoProfile; Neurone; Receptor; Transmembrane.
KW SEQUENCE 906 AA; 101955 MW; 0FB01F24DBD9B0BD CRC64;

Query Match 9.3%; Score 351.5; DB 4; Length 906;
Best Local Similarity 22.2%; Pred. No. 3.7e-18;
Matches 142; Conservative 99; Mismatches 25; Indels 143; Gaps 25;

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Herzog Y., Kalchauer C., Kahane N., Reshef R., Neufeld G.;
RT "Differential expression of neuropilin-1 and neuropilin-2 in arteries
and veins."
RL Mech. Dev. 0:0-0(2002)
DR EMBL; AF417235; AA37723.1; -
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; PS_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; UNKNOWN_2.
DR PROSITE; PS01286; FA58C_2; UNKNOWN_2.
DR PROSITE; PS50060; MAM_2; 1.
DR Receptor.
SQ SEQUENCE 919 AA; 103154 MW; 418AB528D62D59B4 CRC64;

Query Match 9.38; Score 351; DR 13; Length 919;
Best Local Similarity 24.98; Pred. No. 4.1e-18;
Matches 115; Conservative 71; Mismatches 173; Indels 102; Gaps 18;

QY 18 GLALLAVSAPRLQAEELGDCGHLVTYDSTGTMSTKNYPGTYPNHTVCEKTTVPK- 76
DB 9 GFLGCLPGSG---GAETAQPCGGRLNSKADAGYITSPGYNDYPSHQNCWEVIYAPEP 64

QY 77 GKRLILRG-DLDIESQTCADYLLF---TSSDQGYPCGSMYVPEKELLNLTSEYTVR 131
DB 65 NQKIIENFPHEIEKHCKYDFEIRDGDEADLLGKHGNI-APPTIISGSSLYIK 123

QY 132 FESGSHISGRGFLTY-----ASSD-----HPDLITCLERASHYL 166
DB 124 FTSDIAROGAGFSLRYEYKTSGEDCSNFTASNGTISPGFPDKYPHNLDCAVFTIITAKP 183

QY 167 KTE---YSKE-----CPACGRDY-----AGDISG-----NMVDGYRDTSL 198
DB 184 KTEILLHFVLFDELHPDLPQAGEGCKYDMLDIMPQIIPQVGPLIGYCGTKMPSDIRSTTG 243

QY 199 LCKRAIIR--GIIDELGGQISVLQKGISRYEGILANGVLSRQGSLSDKRFLFTSNGCS 256
DB 244 VLSLTFHTDLAVARDGFSQYLLIQOEVPENFQCNVPLGMS--GRISNM----- 291

QY 257 RSLSEFPDGOIRASSSMQSVNESGDQVWSPGOARLDQGPSMASGSSNNHKPREMLEI 316
DB 292 -----QISASSTY-----SDGKWTPOQSRNSDDNGWTPNDVSN---KEYLQV 331

QY 317 DLGKKKKTIGIRTG--STQSNFNFYKSFYVNFKNNSKMKTYKGIYVNEKEVQGSNSN 374
DB 332 DLHFLTVLTAITAGALISRETONGYVRYTYKLEVTNGEDMMWYRHGKNH--KTFQANED 389

QY 375 FRDPYQNNFIPIYARVRYVPQTHQRIALAKVELIGCQIT 415
DB 390 ATEVVLNKHISPLVLRFRVIRPQSHNGIALRLLEYGCRIT 430

Search completed: May 15, 2003, 13:18:46
Job time : 75.8563 secs

PN W0200129088-A1.
 XX 26-APR-2001.
 PD 23-JUN-2000; 2000WO-US17386.
 PF 19-OCT-1999; 99US-0420707.
 XX
 PR
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 PI Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
 XX
 DR WPI: 2001-308477/32.
 DR N-PSDB; AAS00660.
 XX
 PT New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes -
 XX
 PS Claim 8; Fig 1; 263pp; English.
 XX
 CC The sequence represents human TANGO 229 polypeptide. This protein and
 CC similar others exhibit the ability to affect growth, proliferation,
 CC survival, differentiation, activity, morphology, or movement/migration
 CC of, e.g., T cells and cells of the heart, liver, pancreas, placenta,
 CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
 CC blood leukocyte, bone marrow or thymus tissue. They can be used as
 CC modulating agents for regulating cellular processes, thus, the proteins
 CC and their associated nucleic acids can be used to prognosticate, prevent,
 CC diagnose, or treat disorders associated with physiological processes.
 CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,
 CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
 CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
 CC deficit disorder, Crohn's disease, gastroenteritis, goitre,
 CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
 CC muscular dystrophy. Antibodies to disorders such as these can be made by
 CC providing a polypeptide of the invention to an immuno-competent
 CC vertebrate and harvesting blood or serum from the vertebrate.
 CC
 XX
 SQ Sequence 715 AA;
 Query Match 100.0%; Score 3770; DB 22; Length 715;
 Best Local Similarity 100.0%; Pred. No. 1e-306;
 Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYPGARGGALARAAGRLALLAVSAPLRQAELGDCGGHLYTYDGGTMTSKNPG 60
 DB 1 MYPGARGGALARAAGRLALLAVSAPLRQAELGDCGGHLYTYDGGTMTSKNPG 60
 QY 61 TYPNHTVEKITYPKGRLLIRGLDIESQTCASDYLFTSSSDOYGPYCGSMYVKE 120
 DB 61 TYPNHTVEKITYPKGRLLIRGLDIESQTCASDYLFTSSSDOYGPYCGSMYVKE 120
 QY 121 LLLMTSEVTFEESGSHISGRFLITYASSDHPDITCLERASHLTKYEKFCPAGGRD 180
 DB 121 LLLMTSEVTFEESGSHISGRFLITYASSDHPDITCLERASHLTKYEKFCPAGGRD 180
 QY 181 VAGDISGNMVDYRDTSLCKAAIHAGIADLGQISVILQKGISRYEGILANGVLSRD 240
 DB 181 VAGDISGNMVDYRDTSLCKAAIHAGIADLGQISVILQKGISRYEGILANGVLSRD 240
 QY 241 GSLSDKRLFLFNSGCSRSLFEPDQIRASSWSQVNSGQVHMSPOQARLDDGSPSWA 300
 DB 241 GSLSDKRLFLFNSGCSRSLFEPDQIRASSWSQVNSGQVHMSPOQARLDDGSPSWA 300
 QY 301 SGDSNNHKKPREMLEIDGEEKKITGITGTSQSNFNYKSFPMNKNNSKMKYTKG 360
 DB 301 SGDSNNHKKPREMLEIDGEEKKITGITGTSQSNFNYKSFPMNKNNSKMKYTKG 360
 QY 361 IVNNEKVFQGSNFRDVPVQNNFPIPIYARVYRVVPOTHORIALKVELICQITQGNDS 420
 DB 361 IVNNEKVFQGSNFRDVPVQNNFPIPIYARVYRVVPOTHORIALKVELICQITQGNDS 420

QY 421 LVWKTQSSTVSSTKKEDEITTRPIPSEETSGINITTTVAIPLVLLVYVPAKGCIFAA 480
 DB 421 LVWKTQSSTVSSTKKEDEITTRPIPSEETSGINITTTVAIPLVLLVYVPAKGCIFAA 480
 QY 481 RKKKKSGPYSAEAKQDKCKQIKYPPARHQSSEFTSYNEMEMTKLILTSDMADY 540
 DB 481 RKKKKSGPYSAEAKQDKCKQIKYPPARHQSSEFTSYNEMEMTKLILTSDMADY 540
 QY 541 OQPLMIGTGYTRKSGTERPMDTDAEAGVSTDAGGHYDCPQARGHRYALPLAPPEY 600
 DB 541 OQPLMIGTGYTRKSGTERPMDTDAEAGVSTDAGGHYDCPQARGHRYALPLAPPEY 600
 QY 601 ATPYERHVALAHFSAOSGTRVPEPQPKHKSJSSGFSVAVYGAODGYQRPHSQAP 660
 DB 601 ATPYERHVALAHFSAOSGTRVPEPQPKHKSJSSGFSVAVYGAODGYQRPHSQAP 660
 QY 661 ADRGDRKAVSALATESGHPDSQKPPHPTSPSYSAPOCLTLPNOTAMTALL 715
 DB 661 ADRGDRKAVSALATESGHPDSQKPPHPTSPSYSAPOCLTLPNOTAMTALL 715

RESULT 2
 ID A4U00630 standard; Protein: 539 AA.
 AC A4U00630;
 XX 29-APR-2001 (first entry)
 DT
 XX Novel human protein (NHP) sequence #3.
 DE Novel human protein (NHP) sequence #3.
 KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
 KW obesity; high blood pressure; connective tissue disorder; infertility;
 XX NHP-mediated pathway.
 OS Homo sapiens.
 XX
 PN W0200129219-A1.
 PD 26-APR-2001.
 PF 08-OCT-2000; 2000WO-US28798.
 XX
 PR 19-OCT-1999; 99US-0160285.
 PR 18-FEB-2000; 2000US-0183583.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI: 2001-290917/30.
 DR N-PSDB; AAS00615.
 DR
 PT Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 XX
 PS Claim 5; Page 29-30; 33pp; English.
 XX
 CC The sequence represents a novel human protein (NHP) containing a CUB
 CC domain (an extracellular domain). CUB proteins have been associated with
 CC regulating development, modulating cellular processes and preventing
 CC infectious disease. NHP nucleotide sequences are useful for gene therapy
 CC as hybridisation probes for screening libraries and assessing gene
 CC patterns. NHP nucleotide sequences are useful for detecting mutant or
 CC inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These

CC host cells allow for the identification of compounds that bind to NHP
 CC receptors or trigger NHP-mediated pathways.

SQ Sequence 539 AA:

Query Match 74.6%; Score 2812; DB 22; Length 539;
 Best Local Similarity 100.0%; Pred. No. 1.2e-226;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVGAGGALARAAGRLALLAVSAPRLQAEELGDCGHLVYODSGTMTSKNYPG 60
 DB 1 MYVGAGGALARAAGRLALLAVSAPRLQAEELGDCGHLVYODSGTMTSKNYPG 60
 OY 61 TYPNHTVCEKTIIVPKGRKRLRLGLDIESQTCASDYLLFTSSPOYGYCGSMVYKPE 120
 DB 61 TYPNHTVCEKTIIVPKGRKRLRLGLDIESQTCASDYLLFTSSPOYGYCGSMVYKPE 120
 OY 121 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLESASHYLKTEYSKFCPCAGCRD 180
 DB 121 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLESASHYLKTEYSKFCPCAGCRD 180
 OY 181 VAGDISGNMVDGYRDTSLCKAIIHAGIITADELGGQISVLRKGISRYEGLIANGVLSRD 240
 DB 181 VAGDISGNMVDGYRDTSLCKAIIHAGIITADELGGQISVLRKGISRYEGLIANGVLSRD 240
 OY 241 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQVNESGDVHWSPGARLQDOGPSPA 300
 DB 241 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQVNESGDVHWSPGARLQDOGPSPA 300
 OY 301 SGSSNNHNPREFLEIDLGKKRTIGIRTTGSGTOSNPNFYVSKFVNNFKNNKMKTYGK 360
 DB 301 SGSSNNHNPREFLEIDLGKKRTIGIRTTGSGTOSNPNFYVSKFVNNFKNNKMKTYGK 360
 OY 361 IVNNEKEVFGNSNFRDPVONNFIPIVARYVVPVQTHORIALKVELIGCOITGNGDS 420
 DB 361 IVNNEKEVFGNSNFRDPVONNFIPIVARYVVPVQTHORIALKVELIGCOITGNGDS 420
 OY 421 LVWRKTSOSTSVSTKKEDETTIRPISEETSTGINITVAIPLVLVLFVAFMGJFAAF 480
 DB 421 LVWRKTSOSTSVSTKKEDETTIRPISEETSTGINITVAIPLVLVLFVAFMGJFAAF 480
 OY 481 RKKKKGSPYGSAAEQTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLIDLTISDMA 538
 DB 481 RKKKKGSPYGSAAEQTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLIDLTISDMA 538

RESULT 3
 AAU00629

ID AAU00629 standard; Protein; 586 AA.

XX AAU00629;
 XX 29-AUG-2001 (first entry)
 DT XX
 DE Novel human protein (NHP) sequence #2.
 XX
 XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
 KM obesity; high blood pressure; connective tissue disorder; infertility;
 KM NHP-mediated pathway.
 XX
 OS Homo sapiens.
 XX WO200129219-A1.
 XX
 XX 26-APR-2001.
 PD XX
 XX 08-OCT-2000; 2000WO-US28798.
 PF XX
 XX 19-OCT-1999; 99US-0160285.
 PR XX
 XX 18-FEB-2000; 2000US-0183583.
 XX
 PA (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI; 2001-290917/30.
 DR N-PSDB; AAS00614.
 DR XX

PT Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 PS
 XX Claim 2; Page 27-28; 33pp; English.

XX The sequence represents a novel human protein (NHP) containing a CUB
 CC domain (an extracellular domain). CUB proteins have been associated with
 CC regulating development, modulating cellular processes and preventing
 CC infectious disease. NHP nucleotide sequences are useful for gene therapy
 CC of physiological disorders or diseases. NHP oligonucleotides are useful
 CC as hybridization probes for screening libraries and assessing gene
 CC patterns. NHP nucleotide sequences are useful for detecting mutant or
 CC inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These
 CC host cells allow for the identification of compounds that bind to NHP
 CC receptors or trigger NHP-mediated pathways.

SQ Sequence 586 AA:

Query Match 74.6%; Score 2812; DB 22; Length 586;
 Best Local Similarity 100.0%; Pred. No. 1.4e-226;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVGAGGALARAAGRLALLAVSAPRLQAEELGDCGHLVYODSGTMTSKNYPG 60
 DB 48 MYVGAGGALARAAGRLALLAVSAPRLQAEELGDCGHLVYODSGTMTSKNYPG 107
 OY 61 TYPNHTVCEKTIIVPKGRKRLRLGLDIESQTCASDYLLFTSSPOYGYCGSMVYKPE 120
 DB 108 TYPNHTVCEKTIIVPKGRKRLRLGLDIESQTCASDYLLFTSSPOYGYCGSMVYKPE 167
 OY 121 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLESASHYLKTEYSKFCPCAGCRD 180
 DB 168 LLNLTSEVTVRFESGSHISGRGFLTYASSDHPDLITCLESASHYLKTEYSKFCPCAGCRD 227
 OY 181 VAGDISGNMVDGYRDTSLCKAIIHAGIITADELGGQISVLRKGISRYEGLIANGVLSRD 240
 DB 228 VAGDISGNMVDGYRDTSLCKAIIHAGIITADELGGQISVLRKGISRYEGLIANGVLSRD 287
 OY 241 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQVNESGDVHWSPGARLQDOGPSPA 300
 DB 288 GSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSQVNESGDVHWSPGARLQDOGPSPA 347
 OY 301 SGSSNNHNPREFLEIDLGKKRTIGIRTTGSGTOSNPNFYVSKFVNNFKNNKMKTYGK 360
 DB 348 SGSSNNHNPREFLEIDLGKKRTIGIRTTGSGTOSNPNFYVSKFVNNFKNNKMKTYGK 407
 OY 361 IVNNEKEVFGNSNFRDPVONNFIPIVARYVVPVQTHORIALKVELIGCOITGNGDS 420
 DB 408 IVNNEKEVFGNSNFRDPVONNFIPIVARYVVPVQTHORIALKVELIGCOITGNGDS 467
 OY 421 LVWRKTSOSTSVSTKKEDETTIRPISEETSTGINITVAIPLVLVLFVAFMGJFAAF 480
 DB 468 LVWRKTSOSTSVSTKKEDETTIRPISEETSTGINITVAIPLVLVLFVAFMGJFAAF 527
 OY 481 RKKKKGSPYGSAAEQTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLIDLTISDMA 538
 DB 528 RKKKKGSPYGSAAEQTCWKQIKYPAFHOSAEFTISYDNKEKMTOKLIDLTISDMA 585

RESULT 4
 AAU00628

ID AA00628 standard; Protein: 487 AA.
 XX
 AC AA00628;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Novel human protein (NHP) sequence #1.
 XX
 KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
 KW obesity; high blood pressure; connective tissue disorder; infertility;
 KW NHP-mediated pathway.
 XX
 OS Homo sapiens.
 XX
 PN W0200129219-A1.
 XX
 PD 26-APR-2001.
 XX
 PF 08-OCT-2000; 2000WO-0528798.
 XX
 PR 19-OCT-1999; 99US-0160285.
 PR 18-FEB-2000; 2000US-0183583.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX
 DR WPI: 2001-290917/30.
 DR N-PSDB; AAS00613.
 XX
 PT Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 XX
 PS Claim 4; Page 26; 33pp; English.
 XX
 CC The sequence represents a novel human protein (NHP) containing a CUB
 CC domain (an extracellular domain). CUB proteins have been associated with
 CC regulating development, modulating cellular processes and preventing
 CC infectious disease. NHP nucleotide sequences are useful for gene therapy
 CC of physiological disorders or diseases. NHP oligonucleotides are useful
 CC as hybridisation probes for screening libraries and assessing gene
 CC patterns. NHP nucleotide sequences are useful for detecting mutant or
 CC inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These
 CC host cells allow for the identification of compounds that bind to NHP
 CC receptors or trigger NHP-mediated pathways.
 CC
 SO Sequence 487 AA;
 Query Match 67.7%; Score 2551; DB 22; Length 487;
 Best Local Similarity 100.0%; Pred. No. 7.9e-205;
 Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 53 MTSNYPGTYNHRVCEKTIYVPGKRLILRGDIDISQTCASDYLTFTSSDDQGYGC 112
 DB 1 MTSNYPGTYNHRVCEKTIYVPGKRLILRGDIDISQTCASDYLTFTSSDDQGYGC 60
 QY 113 GSMVPRKLLNTSEVYVREFSGSHISGRGLLYVASSDHDLTCTCLERASHYLTETSK 172
 DB 61 GSMVPRKLLNTSEVYVREFSGSHISGRGLLYVASSDHDLTCTCLERASHYLTETSK 120
 QY 173 FCPAGCRVADISGNMVDGRTDLSLCKAIIHAGIINDELJGQISVLRKRGISRYEGIL 232
 DB 121 FCPAGCRVADISGNMVDGRTDLSLCKAIIHAGIINDELJGQISVLRKRGISRYEGIL 180
 QY 233 ANGLYRSGSISDKRFLFTSNGCSRSLSEFPDQIRASSSQSVNESGDQVHWSFGQARL 292

DB 181 ANGLYRSGSISDKRFLFTSNGCSRSLSEFPDQIRASSSQSVNESGDQVHWSFGQARL 240
 QY 293 ODQGPASGDSNNHKKREMLEIDLGKKKKTIGRTGTOSNPNFYKSGFVNNFKNN 352
 DB 241 ODQGPASGDSNNHKKREMLEIDLGKKKKTIGRTGTOSNPNFYKSGFVNNFKNN 300
 QY 353 SKWKTYKGIYVNEEKVFQGSNFRDPVONNFIPIVARYVAVPQTHQRIALKEVIGC 412
 DB 301 SKWKTYKGIYVNEEKVFQGSNFRDPVONNFIPIVARYVAVPQTHQRIALKEVIGC 360
 QY 413 QITGNDSLVWRKTSQSTSVSTKKEDEITIRPISEESTSGINITTVAILPVLVLYEA 472
 DB 361 QITGNDSLVWRKTSQSTSVSTKKEDEITIRPISEESTSGINITTVAILPVLVLYEA 420
 QY 473 GNGJFAFRKKKKKSGPYGSAEAOCTDCWQIKYFPAHQSAEFTISYDNKEWTKIDL 532
 DB 421 GNGJFAFRKKKKKSGPYGSAEAOCTDCWQIKYFPAHQSAEFTISYDNKEWTKIDL 480
 QY 533 ITSDMA 538
 DB 481 ITSDMA 486
 RESULT 5
 AAB19126
 ID AAB19126 standard; Protein: 503 AA.
 XX
 AC AAB19126;
 XX
 DT 19-FEB-2001 (first entry)
 XX
 DE Polypeptide isolated from lymph node stromal cells of fsn -/- mice.
 XX
 KW Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
 KW immune system disorder; cancer; viral infection; HIV infection;
 KW blood vessel growth; tumour necrosis factor disorder; arthritis;
 KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
 KW cardiac failure.
 XX
 OS Mus sp.
 XX
 PN W0200058463-A1.
 PD 05-OCT-2000.
 XX
 PF 18-FEB-2000; 2000WO-NZ00015.
 PR 25-MAR-1999; 99US-0276268.
 PR 26-AUG-1999; 99US-0383586.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX
 PI Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
 PI Murlison JS;
 XX
 DR WPI: 2000-664924/64.
 DR N-PSDB; AAA96736.
 XX
 PT Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
 PT useful for modulating growth of blood cells, for treating inflammatory
 PT and tumour necrosis factor-mediated disorders, cancer and viral
 PT disorders -
 XX
 PS Claim 1; Page 68-69; 75pp; English.
 XX
 CC The present sequence represents a polypeptide sequence which is
 CC isolated from lymph node stromal cells of fsn -/- mice. The
 CC polynucleotides and their polypeptides are useful for treating an
 CC inflammatory disorder, disorder of immune system and cancer selected
 CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
 CC viral disorder, in particular HIV infection and for modulating the
 CC growth of blood vessels. The polypeptides are useful for treating a
 CC tumour necrosis factor (TNF) mediated disorder, such as those selected

CC from arthritis, inflammatory bowel disease and cardiac failure and a
 CC fibroblast growth factor-mediated disorder. It is also useful in assays
 CC to determine biological activity, to raise antibodies, to isolate
 CC corresponding ligands or receptors, to quantify levels of protein or
 CC cognate corresponding ligand or receptors, as anti-inflammatory agents,
 CC and in compositions for the treatment of skin, connective tissue and
 CC immune system diseases. The polynucleotide is useful as marker for
 CC tissue, as a chromosome marker or tags in the identification of a
 CC genetic disorder.

XX Sequence 503 AA;

Query Match 50.7%; Score 1912.5; DB 21; Length 503;
 Best Local Similarity 55.6%; Pred. No. 2,6e-151;
 Matches 390; Conservative 35; Mismatches 74; Indels 203; Gaps 4;

QY 14 AAGGIALALLAVAPRLDAEELGDCGHLVYODSGTMSKRYPCGYTHVTCETIT 73
 DB 5 AGGVSVALPNCAPRLDAEELGDCGHLVYODSGTMSKRYPCGYTHVTCETIT 64
 QY 74 VPKGRLLILRLGLDIESQTCASDYLLFTSSDQYGCYSMTVPKELLNTSEVYRFE 133
 DB 65 VPKGRLLILRLGLDIESQTCASDYLLFTSSDQYGCYSMTVPKELLNTSEVYRFE 124
 QY 134 SGSHISGRGLLYASSDHPDLITCLERASHYLTKEYSKFCPCACRDVAGDISGNMVDGY 193
 DB 125 SGSHISGRGLLYASSDHPDLITCLERASHYLTKEYSKFCPCACRDVAGDISGNMVDGY 184
 QY 194 RDSILCKAIIHGIINDELGGISVLRKISREGLIANGVLSRDSISDRFELTSN 253
 DB 185 RDSILCKAIIHGIINDELGGISVLRKISREGLIANGVLSRDSISDRFELTSN 241
 QY 254 GCSRSISFEEDGOIRASSSQSVNESGDQYHMSFGARLADQGPSMASGDSNNHKKPREW 313
 DB 242 ----- 241
 QY 314 LEIDLGKKKRTIGRTGTSQSNFNYKSFVNNKNNNSKMYKGIYNNKRVFGNS 373
 DB 242 ----- 241
 QY 374 NFRDPVQNNFPIPVAVYVYVQTHQRIALNVELIGCQITQGNDSLVRKTSQTSVS 433
 DB 242 ----- 241
 QY 434 TKKEDETTPRIPSEETSTGINTTVAIPVLVLFVAGMGIAPARKKKKKSPYGS 493
 DB 242 ----- 285
 QY 494 EAOKTDCWKQIKPPARHOSAEFTISTDNKEKMTOKLIDITSMDADYQOPLMIGTVTR 553
 DB 286 DAOKTDCWKQIKPPARHOSAEFTISTDNKEKMTOKLIDITSMDADYQOPLMIGTVTR 345
 QY 554 KGSFPRMDADAEAGYSTAGGHYDCROBAGREVALPLAPPEEATPVERHVLRAH 613
 DB 346 KGSFPRMDADAEAGYSTAGGHYDCROBAGREVALPLAPPEEATPVERHVLRAH 405
 QY 614 TFSASQSYRVPGDPGKHSLSSGSGFSPVAGVAGNODGYORPHASQPADRDPKAVSA 673
 DB 406 TFSASQSYRVPGDPGKHSLSSGSGFSPVAGVAGNODGYORPHASQPADRDPKAVSA 464
 QY 674 LATESGPPDSQKPTTHGTSDSISAPDCITPLNQTMTALL 715
 DB 465 L-DSRDPASQSQWT-SGGDGYSAFPRNGLAFLNQTMTALL 503

RESULT 6
 AAE22715
 ID AAE22715 standard; Protein; 398 AA.

XX AAE22715;
 AC
 XX
 DT 09-AUG-2002 (first entry)
 XX

DE Human neuropilin-Hy1 protein.

XX Human: neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoa; di-
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnerability; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virocid; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.

XX Homo sapiens.

XX WO200222815-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28488.

XX 11-SEP-2001; 2000US-0659671.

XX 06-SEP-2001; 2001US-317902P.

XX (HYSE-) HYSEQ INC.

XX Tang YT;

XX WPI: 2002-393966/42.

XX N-PSDB; AAD35992.

PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects

PS Claim 3; Page 123-125; 152pp; English.

CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy1
 CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

SO Sequence 398 AA;

Query Match 37.4%; Score 1409.5; DB 23; Length 398;
 Best Local Similarity 72.8%; Pred. No. 2,6e-109;

Matches 289; Conservative 11; Mismatches 26; Indels 71; Gaps 7;

QY 38 GDGCGHLYTYDDSGMTSKNTPGTYPNHTVCERTITVPKGRKLLIRLGDLDIESQTCASD 97
 DB 4 GDGCGHLYTYDDSGMTSKNTPGTYPNHTVCERTITVPKGRKLLIRLGDLDIESQTCASD 63
 QY 98 YLLFTSSDQGYPGYSGMTVPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDLIT 157
 DB 64 YLLFTSSDQGYPGYSGMTVPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDLIT 121
 QY 158 CLERSHLYKTEYSK-FCPACGRDVAAGDISGMMVDGYRDTSLCKRAIHAGIADDELGGQ 216
 DB 122 QGDRSEKTLDOQSRTFLATGTFVKDSFS--TDG---TSLCKRAIHAGIADDELGGQ 175
 QY 217 ISVLORKGISREGLANGVLS-----ELFTSNCGSRSLSFEPDGOIRASSSMQS 238
 DB 176 ISVLORKGISREGLANGVLSREFEIRFQELSSVLFYSGMNTVHAVIELMPPMIYWH 235
 QY 239 ----RDGSLSDKR-----ELFTSNCGSRSLSFEPDGOIRASSSMQS 275
 DB 236 SGTREGSIAAEEEGVPLLYLVYIQKQELVODLVAVATGCSRSLSFEPDGOIRASSSMQS 295
 QY 276 VNESGDQVHWSFGQARLDQGPMSASGDSNNHKPREMLEIDLGKKKITGIRTTGSTQS 335
 DB 296 VNESGDQVHWSFGQARLDQGPMSASGDSNNHKPREMLEIDLGKKKITGIRTTGSTQS 355
 QY 336 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEVFOGN 372
 DB 356 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEV-RCN 391

RESULT 7
 AAU79459 standard; Protein: 398 AA.

AAU79459:
 15-JUL-2002 (first entry)
 Human Neuropilin-Hy1.

Human: neuropilin-Hy1; chromosome 6q21; neuronal growth;
 nerve regeneration; neurodegenerative disease; learning disorder;
 memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
 organ growth; nervous system lesion; cancer; cell proliferation;
 cell differentiation; stem cell growth factor activity;
 Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
 haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
 platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
 reperfusion; food supplement; DNA microarray.

Homo sapiens.
 WO200222780-A2.
 21-MAR-2002.
 11-SEP-2001; 2001WO-US285390.
 11-SEP-2000; 2000US-0659671.
 06-SEP-2001; 2001US-0659671.
 (TANG/) TANG T Y.
 Tang TY;
 WPI: 2002-351881/38.
 N-PSDB: ABA49565.

New neuropilin-like polypeptides for diagnosing, preventing and
 treating neurological conditions and disorders, cancers, and for
 inducing angiogenesis and neovascularisation

PS Claim 3; Page 118-120; 144pp; English.

XX The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC mapping genetic neuronal defects and degenerative diseases like
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischemic lesions, infectious lesions,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for
 CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopaenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hy1.

Sequence 398 AA:
 SQ
 37.48; Score 1409.5; DB 23; Length 398;
 Best Local Similarity 72.8%; Pred. No. 2.6e-109;
 Matches 289; Conservative 11; Mismatches 26; Indels 71; Gaps 7;

QY 38 GDGCGHLYTYDDSGMTSKNTPGTYPNHTVCERTITVPKGRKLLIRLGDLDIESQTCASD 97
 DB 4 GDGCGHLYTYDDSGMTSKNTPGTYPNHTVCERTITVPKGRKLLIRLGDLDIESQTCASD 63
 QY 98 YLLFTSSDQGYPGYSGMTVPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDLIT 157
 DB 64 YLLFTSSDQGYPGYSGMTVPKRELLNTSEVTVRESGSHISGRGFLTYASSDHPDLIT 121
 QY 158 CLERSHLYKTEYSK-FCPACGRDVAAGDISGMMVDGYRDTSLCKRAIHAGIADDELGGQ 216
 DB 122 QGDRSEKTLDOQSRTFLATGTFVKDSFS--TDG---TSLCKRAIHAGIADDELGGQ 175
 QY 217 ISVLORKGISREGLANGVLS-----ELFTSNCGSRSLSFEPDGOIRASSSMQS 238
 DB 176 ISVLORKGISREGLANGVLSREFEIRFQELSSVLFYSGMNTVHAVIELMPPMIYWH 235
 QY 239 ----RDGSLSDKR-----ELFTSNCGSRSLSFEPDGOIRASSSMQS 275
 DB 236 SGTREGSIAAEEEGVPLLYLVYIQKQELVODLVAVATGCSRSLSFEPDGOIRASSSMQS 295
 QY 276 VNESGDQVHWSFGQARLDQGPMSASGDSNNHKPREMLEIDLGKKKITGIRTTGSTQS 335
 DB 296 VNESGDQVHWSFGQARLDQGPMSASGDSNNHKPREMLEIDLGKKKITGIRTTGSTQS 355
 QY 336 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEVFOGN 372
 DB 356 NFNFYKSFVNFKNNSKMKTYKGIYVNEKEV-RCN 391

RESULT 8
 AAU70539 standard; Protein: 669 AA.

AB97386	AB97386 standard; Protein; 385 AA.
AC	AB97386;
DT	27-JUN-2002 (first entry)
DE	Novel human protein SEQ ID NO: 654.
XX	Human; anti-nausea; vulnary; anti-inflammatory; immunomodulator;
XX	anti-inflamm; cerebroprotective; cytosolic; rheumatic; gene therapy;
XX	neuroprotective; antiparkinsonian; protein therapy; EST;
XX	expressed sequence tag.
OS	Homo sapiens.
PN	WO200222660-A2.
PD	21-MAR-2002.
XX	10-SEP-2001; 2001WO-US26015.
PF	11-SEP-2000; 2000US-0659671.
PR	(HYSE-) HYSEQ INC.
PA	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI	Xue AJ, Yang Y, Wehrman T, Drmanac RT;
PI	WPI: 2002-292408/33.
XX	N-PSDB: ABN32572.
DR	An isolated polynucleotide for treating diseases associated with its
DR	encoded polypeptide such as cancer and multiple sclerosis -
XX	Claim 20; SEQ ID NO 654; 509pp; English.
XX	The present invention provides the protein and coding sequences of 444
CC	novel human proteins. These were isolated from expressed sequences tags
CC	(ESTs). They can be used to stimulate cell growth, to regulate
CC	haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC	e.g. in burn treatment, to regulate the immune system e.g. to treat
CC	multiple sclerosis, to regulate activin or inhibit e.g. to treat
CC	infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC	stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC	Parkinson's disease. The present sequence is a protein of the invention.
XX	Sequence 385 AA;
XX	Query Match 18.4%; Score 692.5; DB 23; Length 385;
XX	Best Local Similarity 41.8%; Pred. NO. 3e-49; Mismatches 131; Indels 31; Gaps 13.
XX	Matches 163; Conservative 65; Mismatches 131; Indels 31; Gaps 13.
XX	20 LALLAVAPRLQAELDGGCGHVVYODSGTMSKNYPGTYPMHVCERTYTPKGR 79
XX	5 LLLLVLLVLLLEDAGAGGCGGCHVVLGPESTLSINVPQYIPNSTVCWEWEIRVMGER 64
XX	80 LILRLGDDIE-SQVCADYLLFTS-----SSDQYPCG-SMTVREKLLANTSEVTPRE 133
XX	65 VRINKGDDIEDSDCHNRYLNTYVIGVSRLEIKYCGLGLOMHSIESKNEETLLPM 124
XX	134 GSHHSGRGFLTYVSSPHDLITCERASHLTKTEYSKFCPCAGCRDVAADISGMNVQGY 193
XX	125 SGHNSGGGFLASIVDKODLITCLIDRASNLFEPEFSKYCAAGCLIPPAEISGITPHGY 184
XX	194 RDTSLAKRAAIIAGIADLGGQISVLQKGISREGLIANGVLNRDGLSDKRRLETSN 253
XX	185 RDSPLCAAGVAGVSNLTGGQISVLTSGKIPYESSLANNVTSVGHLSTSLTFKTS 244
XX	254 GCSRLSTFEPDQQLTASS-----WQSVNMSGQVHMSPEQARLQDQGPFW-ASGDS 305
XX	245 GCGYGLMGESGD-RGSSNNSTITVLEW--TDHTTGGENSMKPKKSOAE--KTMTALGARA 298

Qy	306	NNHPRLEHLEIDGEEKKIGIRGTSTGTO-SNFNFVKS---FVNMKKNNSKMTYGI	361
Db	299	TDE--YQWLQIDLNKEKKITGITITTSSTWKTITMCLPSTECTVMGRNG---	LCSTSL 352
Qy	362	VNNEKVFQGNNSN-FRDPVQNNFPIVAR	390
Db	353	VVEQDKIFQGNKRITIRTMVNNFLPPIAR	382
RESULT 13			
ID	AAE22721	AAE22721 standard; Protein; 365 AA.	
AC	AAE22721:		
XX	09-AUG-2002	(first entry)	
DT			
DE		Human neuropilin-Hy2 mature protein.	
XX			
KW		Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;	
KW		neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;	
KW		thrombocytopenia; memory; platelet; plastic anaemia; antiinflammatory;	
KW		ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;	
KW		wound healing; tissue repair; Parkinson's disease; Huntington's disease;	
KW		amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;	
KW		cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;	
KW		human immunodeficiency virus; HIV; autoimmune disorder; dermatological;	
KW		systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;	
KW		autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;	
KW		myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;	
KW		neurologic; neuroprotective; vulnery; anticonvulsant; antiparasitic;	
KW		cerebroprotective; tranquiliser; virucide; antibacterial; cytostatic;	
KW		immunosuppressive.	
XX			
XX			
OS		Homo sapiens.	
XX			
PN	WO200222815-A1.		
XX			
PD	21-MAR-2002.		
XX			
PE	12-SEP-2001; 2001MO-US28488.		
XX			
XX	11-SEP-2000; 2000US-0659671.		
PR	06-SEP-2001; 2001US-317902P.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
P1	Tang YT;		
XX			
DR	WPI: 2002-393966/42.		
XX			
PT	Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides		
PT	useful for treating neurodegenerative diseases e.g. Alzheimer's		
PT	disease, and for diagnosing and mapping genetic neuronal defects		
XX			
PS	Disclosure; Page 131-132; 152pp; English.		
XX			
CC		The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and	
CC		neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-	
CC		like polypeptides and polynucleotides are useful in modulating neuronal	
CC		growth regenerative capacity, treating neurodegenerative diseases,	
CC		diagnosing and mapping genetic neuronal defects and degenerative diseases	
CC		like Alzheimer's disease and for treating learning and memory disorders.	
CC		They are also useful for inducing angiogenesis, neovascularisation, as	
CC		well as organ growth and development e.g. heart and other tissues.	
CC		Antagonists of neuropilin-like polypeptides are useful for treating	
CC		cancers and other malignant diseases. Neuropilin is used to treat	
CC		platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal	
CC		nocturnal haemoglobinuria and is used in nerve tissue growth or	
CC		regeneration, in wound healing, tissue repair and replacement and in	
CC		healing of bones, incisions and ulcers. Compositions comprising the	
CC		sequences of the invention are useful for treating diseases of peripheral	

ID AAM96308 standard; Protein: 925 AA.
 AC AAM96308;
 XX 28-JUN-1999 (first entry)
 XX Neuropllin-2.
 DE Neuropllin-2.
 XX Neuropllin; neuropllin-2; semaphorin; transmembrane protein; axon;
 KW neurone; development; cell growth; immune response;
 KW viral pathogenesis; treatment; disease; graft rejection;
 KW viral disease; oncological disease; cancer; screening; probes.
 XX Rattus rattus.
 OS
 XX MO9904263-A1.
 XX 28-JAN-1999.
 XX 17-JUL-1998; 98WO-US14632.
 XX 17-JUL-1997; 97US-0052762.
 PR (UyJo) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 PA
 PI Glincy DD, Kolodkin AL;
 DR WPI; 1999-132446/1.
 DR N-PSDB; AAX08417.
 XX
 PT New isolated semaphorin receptor, neuropllin-2 - used to develop
 PT products for the diagnosis and treatment of neurological,
 PT immunological, oncological and viral diseases
 PS
 PS Claim 13; Page 77-79; 88pp; English.
 XX
 CC The neuropllins are type I transmembrane proteins and act as
 CC semaphorin III (Sema III) receptors. The semaphorins have been
 CC shown to function in repulsive axon guidance. Sema III is a
 CC secreted protein that in vitro causes neuronal growth cone
 CC collapse and chemorepulsion on neurites and is required in vivo for
 CC correct sensory afferent innervation and other aspects of
 CC development. Agents which inhibit or enhance the interaction of a
 CC semaphorin and a neuropllin can be potent modulators of nerve cell
 CC growth, immune responsiveness, and viral pathogenesis, and can be
 CC used in the treatment and diagnosis of neurological disease,
 CC neuro-regeneration, immune modulation including hypersensitivity and
 CC graft-rejection, and diagnosis and treatment of viral and oncological
 CC infection/diseases. The neuropllins, neuropllin-encoding nucleic
 CC acids, and unique portions also are useful in screening chemical
 CC libraries for regulators of semaphorin-mediated cell activity, and in
 CC genetic mapping as probes for related genes, as diagnostic reagents
 CC for genetic, neurological, immunological and oncological disease.
 CC
 SO Sequence 925 AA;

Query Match 9.6%; Score 362.5; DB 20; Length 925;
 Best local Similarity 22.3%; Pred. No. 5.2e-21;
 Matches 143; Conservative 100; Mismatches 246; Indels 151; Gaps 24;

OY 41 CGHLVYDDSGTMTSKNPGYTPNTVEKITVPR-GKRLIRLG-DLDESQCASDY 98
 DB 28 CGGRINSKADGYTSPGYDPSHONCEWVYAPENQKIVLNPNPFEIEKHCKYDF 87
 OY 99 LIF-----TSSDOQGYPGYSGMTVPKELLANTSEVTRFESGSHISGRFLLTY-----AS 149
 DB 88 IEIRGDSESDLDLKGHCNT-APPTIISSGSVLYIKFTSDYAROGAGFSLRYELFKTGS 146
 OY 150 SD-----HPDLIFC-----LERASHYLKTEYSKF----- 173
 DB 147 EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPRMEIILQFLTFDLHDPLQVGE 206
 OY 174 -CPAGCRVADIS--GMVNDGYRDTSLCKRAIHAIGIADLGGQISVLRKGISRY-- 228

DB 207 DCKYDMLDIPWGIPIHVGPIGKTCGTPSKLRSSTGILSLTFHTDMAVADGFSARYYL 266
 OY 229 -----EGILANGVYSRD-GSLSDKRFLFTSNGCSRSLSPFDDGIRASSSQSVNEGSD 281
 DB 267 VHQEPPEFQCNAPLGMESGRIANE-----QISASSTF-----S 300
 OY 282 QVHWSPGQARLQDQGPWSASGDDSSNNHPRFMLEIDGKKRTIGRTG--STQSNFNF 339
 DB 301 DGRMTPOQSRLHGDNDNGWTPNVDSN---KEYLDVDFLMLPAIATGALISRETQCY 356
 OY 340 YKSPVAMFNKNNKWKTKYKGVNNEEKVFOGNSFRPQONNFPPIVAVYRVPTW 399
 DB 357 YKSYKLEVSTNGEDMVMYRGRKNH--KVFOANNDATELVNKLHTPLTRIRLPQTW 414
 OY 400 HORIALKVELIGCOITGQNDSLVWRKTS---QSTVSSTKKEDETTRPI----- 445
 DB 415 HIGIALRIELFGCRKTDAPCSNMLGSLADTQISASTREYIMSSAARLYSSRSQW 474
 OY 446 -----PSEETSTGINTTVAIPVLVLLVYVFGMGIFA---AFRRKKRKGSPYGA 493
 DB 475 FPRNPQAPGGEW-LQVDLGPPTVKGVILQARGGDSITAMEARAFYRKFVSYSLNGK 533
 OY 494 EAQKTDCKKQIKYPPARQSA-EFTISYD-----NEKEMTOKL 530
 DB 534 D-----WEYIADPPTQPKLEEGNMAYDTPDIRFEFVPAQYRVYPERMSPAGIGRL 587
 OY 531 DLITSDMADYOQPL-MIGTGVTRKSGTFPRMDTDAEEAG 569
 DB 588 EVLGCDWTDNSKPTVETLGPYKSEETTPPYMDEBATECG 627

Search completed: May 15, 2003, 13:20:42
 Job time : 68.5731 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:15:30 ; Search time 24.0681 Seconds
(without alignments)
2865.294 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MWPARGGALARAAGRL.....YSAPRDLPLNQTALL 715Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications.AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3770	100.0	715	US-10-003-132-2	Sequence 2, Appl1
2	3770	100.0	715	US-09-759-130B-73	Sequence 73, Appl1
3	3610	95.8	681	US-09-759-130B-75	Sequence 75, Appl1
4	2229	59.1	421	US-09-759-130B-76	Sequence 76, Appl1
5	1924.5	51.0	503	US-10-003-132-4	Sequence 4, Appl1
6	1912.5	50.7	503	US-09-823-038A-51	Sequence 51, Appl1
7	1650	44.8	458	US-10-003-132-6	Sequence 6, Appl1
8	1265	33.6	235	US-09-759-130B-78	Sequence 78, Appl1
9	1042	27.6	729	US-10-060-830-3	Sequence 3, Appl1
10	873.5	23.2	653	US-10-060-830-1114	Sequence 1114, Appl1
11	351.5	9.3	931	US-10-104-440-4	Sequence 4, Appl1
12	335	8.9	923	US-10-104-440-2	Sequence 2, Appl1
13	317	8.4	2224	US-10-115-563-14	Sequence 14, Appl1
14	306.5	8.1	1431	US-10-095-718-4	Sequence 4, Appl1
15	299	7.9	343	US-10-190-593-2	Sequence 2, Appl1
16	299	7.9	387	US-10-190-593-4	Sequence 4, Appl1
17	285.5	7.6	1438	US-10-225-900-1	Sequence 1, Appl1
18	285.5	7.6	1438	US-10-006-091-1	Sequence 1, Appl1
19	285.5	7.6	1438	US-10-047-257-1	Sequence 1, Appl1

20	285.5	7.6	1471	12	US-10-095-718-2	Sequence 2, Appl1
21	285.5	7.6	2332	9	US-09-957-641-2	Sequence 2, Appl1
22	285.5	7.6	2332	9	US-10-187-319-2	Sequence 2, Appl1
23	285.5	7.6	2332	9	US-10-132-829-4	Sequence 4, Appl1
24	276.5	7.3	2319	9	US-10-187-319-6	GENERAL INFORNA
25	271.5	7.2	1443	9	US-10-187-319-39	Sequence 39, Appl1
26	271.5	7.2	2133	9	US-10-187-319-37	Sequence 37, Appl1
27	226	6.0	1128	9	US-09-996-015-10	Sequence 10, Appl1
28	226	6.0	1128	9	US-10-238-876-8	Sequence 8, Appl1
29	217	5.8	845	9	US-09-996-015-11	Sequence 11, Appl1
30	217	5.8	1158	9	US-10-238-876-2	Sequence 2, Appl1
31	216.5	5.7	719	9	US-09-996-015-9	Sequence 9, Appl1
32	185.5	4.9	250	10	US-09-770-643A-8	Sequence 8, Appl1
33	185.5	4.9	279	10	US-09-770-643A-10	Sequence 10, Appl1
34	185.5	4.9	534	10	US-09-770-643A-14	Sequence 14, Appl1
35	185.5	4.9	582	10	US-09-770-643A-12	Sequence 12, Appl1
36	185.5	4.9	697	10	US-09-770-643A-18	Sequence 18, Appl1
37	185.5	4.9	745	10	US-09-770-643A-16	Sequence 16, Appl1
38	185.5	4.9	791	10	US-09-770-643A-20	Sequence 20, Appl1
39	185.5	4.9	839	10	US-09-770-643A-22	Sequence 22, Appl1
40	185.5	4.9	1259	10	US-09-770-643A-4	Sequence 4, Appl1
41	185.5	4.9	1307	10	US-09-770-643A-2	Sequence 2, Appl1
42	184.5	4.9	1298	10	US-09-770-643A-24	Sequence 24, Appl1
43	184.5	4.9	1298	10	US-09-770-643A-28	Sequence 28, Appl1
44	182.5	4.8	764	10	US-09-900-518A-2	Sequence 2, Appl1
45	180.5	4.8	686	10	US-09-874-198-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Shoenaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

Query Match 100.0%; Score 3770; DB 9; Length 715;
Best Local Similarity 100.0%; Pred. No. 8e-284;
Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MWPARGGALARAAGRLALILAVSAPRLAEEIGDCGHLVYQDSGTMTSKYPG 60
DB 1 MWPARGGALARAAGRLALILAVSAPRLAEEIGDCGHLVYQDSGTMTSKYPG 60
QY 61 TYPNHTVCERTITVPKRRILRLRGDIDISQTCASDYLLFTSSSDQYGPYCSMTYPK 120
DB 61 TYPNHTVCERTITVPKRRILRLRGDIDISQTCASDYLLFTSSSDQYGPYCSMTYPK 120
QY 121 LLNTSEVTVRFESGSHISGRGLTYASSDHPDLITCLERASHYLTETSKPCPACRD 180
DB 121 LLNTSEVTVRFESGSHISGRGLTYASSDHPDLITCLERASHYLTETSKPCPACRD 180
QY 181 VAGDISNMVDGYRDSILCKAIIHAGITIDELGGQISVLRGKISREKIIANGVLSRD 240
DB 181 VAGDISNMVDGYRDSILCKAIIHAGITIDELGGQISVLRGKISREKIIANGVLSRD 240

```

QY 241 GSLSDKRLFTSNGCSRSLSEFPDQIRASSSMOSVNSGQVHMSPGQARLDQGP5MA 300
D 241 GSLSDKRLFTSNGCSRSLSEFPDQIRASSSMOSVNSGQVHMSPGQARLDQGP5MA 300
QY 301 SGDSNNHKKPREMLEIDEGEKKKITGITRTGTSQSNFNFYKSFVNMKNNSKMKTKYG 360
D 301 SGDSNNHKKPREMLEIDEGEKKKITGITRTGTSQSNFNFYKSFVNMKNNSKMKTKYG 360
QY 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPPQTHORIALKVELICQITQGNDS 420
D 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPPQTHORIALKVELICQITQGNDS 420
QY 421 LVNRTSGTSVSTRKKEDETTTRPIPSEETSGINITTVALPIVLVLYVFRAGMIFAF 480
D 421 LVNRTSGTSVSTRKKEDETTTRPIPSEETSGINITTVALPIVLVLYVFRAGMIFAF 480
QY 481 RKKKKGSPYSAEAKTDCKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
D 481 RKKKKGSPYSAEAKTDCKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
QY 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHYDCQORAGRHEYALPLAPEPEY 600
D 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHYDCQORAGRHEYALPLAPEPEY 600
QY 601 ATPYERHVLAAHTFSAQSGRYRVPQPGHKSLSGGFSPVAGVAGDGDYORPHSAP 660
D 601 ATPYERHVLAAHTFSAQSGRYRVPQPGHKSLSGGFSPVAGVAGDGDYORPHSAP 660
QY 661 ADRGYDRKAVSALATESGHPDSOKPPTHPTSDSYSAPRDCLPLNOTANTALL 715
D 661 ADRGYDRKAVSALATESGHPDSOKPPTHPTSDSYSAPRDCLPLNOTANTALL 715

RESULT 2
US-09-759-130B-73
: Sequence 73, Application US/09759130B
: Publication No. US20030022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D
: APPLICANT: Barnes, Thomas S
: APPLICANT: Kirst, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wighton, Nicolas
: APPLICANT: Goodearl, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OR INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OR INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OR INVENTION: USES.
: FILE REFERENCE: MP100-5350NIM
: CURRENT APPLICATION NUMBER: US/09/759, 130B
: PRIOR FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479, 249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559, 497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578, 063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333, 159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596, 194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342, 364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608, 452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393, 996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602, 871

```

```

: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420, 707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 73
: LENGTH: 715
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-759-130B-73

Query Match      100.0%; Score 3770; DB 9; Length 715;
Best Local Similarity 100.0%; Pred. No. 8e-284;
Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYPGARGGALARAAGRLALLAVASAPRLIOAEELDGGCHLVYQDSTMTSKNPG 60
D 1 MYPGARGGALARAAGRLALLAVASAPRLIOAEELDGGCHLVYQDSTMTSKNPG 60
QY 61 TYPNHTVEKTIYPPKGRLLRLGDDLDESQTCASDYLFTSSSDYGPYCGSKTYPKE 120
D 61 TYPNHTVEKTIYPPKGRLLRLGDDLDESQTCASDYLFTSSSDYGPYCGSKTYPKE 120
QY 121 LLNLTSEVTFEESGSHISGRGFLITVASSDHPDLITCERASHYLTKEYSKPCAGCD 180
D 121 LLNLTSEVTFEESGSHISGRGFLITVASSDHPDLITCERASHYLTKEYSKPCAGCD 180
QY 181 VAGDISGMVNDYRDTSLCKAAIHAGIADLGGQISVLORKGISREYEGILANGVLSRD 240
D 181 VAGDISGMVNDYRDTSLCKAAIHAGIADLGGQISVLORKGISREYEGILANGVLSRD 240
QY 241 GSLSDKRLFTSNGCSRSLSEFPDQIRASSSMOSVNSGQVHMSPGQARLDQGP5MA 300
D 241 GSLSDKRLFTSNGCSRSLSEFPDQIRASSSMOSVNSGQVHMSPGQARLDQGP5MA 300
QY 301 SGDSNNHKKPREMLEIDEGEKKKITGITRTGTSQSNFNFYKSFVNMKNNSKMKTKYG 360
D 301 SGDSNNHKKPREMLEIDEGEKKKITGITRTGTSQSNFNFYKSFVNMKNNSKMKTKYG 360
QY 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPPQTHORIALKVELICQITQGNDS 420
D 361 IVNNEEKVFGQNSNFRDPVQNNFIPPIVARYRVVPPQTHORIALKVELICQITQGNDS 420
QY 421 LVNRTSGTSVSTRKKEDETTTRPIPSEETSGINITTVALPIVLVLYVFRAGMIFAF 480
D 421 LVNRTSGTSVSTRKKEDETTTRPIPSEETSGINITTVALPIVLVLYVFRAGMIFAF 480
QY 481 RKKKKGSPYSAEAKTDCKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
D 481 RKKKKGSPYSAEAKTDCKQIKYPPARHQAFTISYNEKEMTKLDLITSDMADY 540
QY 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHYDCQORAGRHEYALPLAPEPEY 600
D 541 QOPLMIGTGTTRKGSFRPMDTDAEAGVSTDAGHYDCQORAGRHEYALPLAPEPEY 600
QY 601 ATPYERHVLAAHTFSAQSGRYRVPQPGHKSLSGGFSPVAGVAGDGDYORPHSAP 660
D 601 ATPYERHVLAAHTFSAQSGRYRVPQPGHKSLSGGFSPVAGVAGDGDYORPHSAP 660
QY 661 ADRGYDRKAVSALATESGHPDSOKPPTHPTSDSYSAPRDCLPLNOTANTALL 715
D 661 ADRGYDRKAVSALATESGHPDSOKPPTHPTSDSYSAPRDCLPLNOTANTALL 715

RESULT 3
US-09-759-130B-75
: Sequence 75, Application US/09759130B
: Publication No. US20030022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D

```

```

: APPLICANT: Barnes, Thomas S
: APPLICANT: Kist, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wrighton, Nicolas
: APPLICANT: Goodheart, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OF INVENTION: USES
: FILE REFERENCE: MP100-5350NM14
: CURRENT APPLICATION NUMBER: US/09/759,130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479,249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559,497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578,063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333,159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596,194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342,364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608,452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393,996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602,871
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420,707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 75
: LENGTH: 681
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-759-130B-75

Query Match          95.8%; Score 3610; DB 9; Length 681;
Best Local Similarity 100.0%; Pred. No. 1.9e-271;
Matches 681; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 EELGDCGHLVYODSGMTSKNYPCTYNNHYCEKTIYVPGKRLILRLGDDISSQC 94
DB 1 EELGDCGHLVYODSGMTSKNYPCTYNNHYCEKTIYVPGKRLILRLGDDISSQC 60
QY 95 ASDYLLFTSSDOYGPYCGSMYVPEKLLNTSEVTVFESSGSHISGRGFLTYAASDHPD 154
DB 61 ASDYLLFTSSDOYGPYCGSMYVPEKLLNTSEVTVFESSGSHISGRGFLTYAASDHPD 120
QY 155 LITCLERASHYLTETXSKFCPCAGCDVAGDISGNMVDGYRDTSLCKAAIHAGIADDELG 214
DB 121 LITCLERASHYLTETXSKFCPCAGCDVAGDISGNMVDGYRDTSLCKAAIHAGIADDELG 180
QY 215 GQISVQARGISRYEGIIANGVLSRDGSLSDKRFLETSNGCSLSLFEEDGQIRASSMQ 274
DB 181 GQISVQARGISRYEGIIANGVLSRDGSLSDKRFLETSNGCSLSLFEEDGQIRASSMQ 240
QY 275 SYNESGDQVHMSPGARLDQGPSPNAGSDSSNNHKKPREMLEIDLGKKKITGIRTGSTG 334
DB 241 SYNESGDQVHMSPGARLDQGPSPNAGSDSSNNHKKPREMLEIDLGKKKITGIRTGSTG 300
QY 335 SNFNFYVSKFVNMFKNNNSKMTYKGIYVNEEKVFQGSNFRDPVONNFIPIVARYRV 394
DB 301 SNFNFYVSKFVNMFKNNNSKMTYKGIYVNEEKVFQGSNFRDPVONNFIPIVARYRV 360
QY 395 VPQTHQRIALVELLIGCITIGGNDSLVWRKTSQSTSVSTKKEDEITIRPISEETSTGI 454
DB 361 VPQTHQRIALVELLIGCITIGGNDSLVWRKTSQSTSVSTKKEDEITIRPISEETSTGI 420

```

```

QY 455 NITTVAIPLVLVLYVFAAGGIFAARFKKKKSPYGSABQKTDCKWKOIKYPFAHHQSA 514
DB 421 NITTVAIPLVLVLYVFAAGGIFAARFKKKKSPYGSABQKTDCKWKOIKYPFAHHQSA 480
QY 515 EFTISYDNEKEKMTOKLDLITSDMADYQOPLMIGTVYTRKSTFFRPMDDAEAGVSTDA 574
DB 481 EFTISYDNEKEKMTOKLDLITSDMADYQOPLMIGTVYTRKSTFFRPMDDAEAGVSTDA 540
QY 575 GGHYDCPQARGREHYALPLAPPEPEYATPIVERHYLRAHFFSAQSGYRVGCPQGRKHS 634
DB 541 GGHYDCPQARGREHYALPLAPPEPEYATPIVERHYLRAHFFSAQSGYRVGCPQGRKHS 600
QY 635 SSGFSFVAVGAGDGDYQPHSAOPADRGYDRPKAVSALATESGHPDSOKPTHPGTSD 694
DB 601 SSGFSFVAVGAGDGDYQPHSAOPADRGYDRPKAVSALATESGHPDSOKPTHPGTSD 660
QY 695 SYSAPRDCITPLNOTAMTALL 715
DB 661 SYSAPRDCITPLNOTAMTALL 681

```

```

RESULT 4
US-09-759-130B-76
: Sequence 76, Application US/09759130B
: Publication No. US2003002279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: APPLICANT: McCarthy, Sean A
: APPLICANT: Fraser, Christopher C
: APPLICANT: Sharp, John D
: APPLICANT: Barnes, Thomas S
: APPLICANT: Kist, Susan J
: APPLICANT: Mackay, Charles R
: APPLICANT: Myers, Paul S
: APPLICANT: Leiby, Kevin R
: APPLICANT: Wrighton, Nicolas
: APPLICANT: Goodheart, Andrew
: APPLICANT: Holtzman, Douglas A
: TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
: TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
: TITLE OF INVENTION: USES
: FILE REFERENCE: MP100-5350NM14
: CURRENT APPLICATION NUMBER: US/09/759,130B
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: US 09/479,249
: PRIOR FILING DATE: 2000-01-07
: PRIOR APPLICATION NUMBER: US 09/559,497
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 09/578,063
: PRIOR FILING DATE: 2000-05-24
: PRIOR APPLICATION NUMBER: US 09/333,159
: PRIOR FILING DATE: 1999-06-14
: PRIOR APPLICATION NUMBER: US 09/596,194
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 09/342,364
: PRIOR FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: US 09/608,452
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/393,996
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: US 09/602,871
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 09/420,707
: PRIOR FILING DATE: 1999-10-19
: NUMBER OF SEQ ID NOS: 460
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 76
: LENGTH: 421
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-759-130B-76

```

Query Match 59.1%; Score 2229; DB 9; Length 421;
 Best Local Similarity 100.0%; Pred. No. 1.1e-164;
 Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 EELDGGCHLVTYODSGMTSKNPGTYPNHTVCEKITVTKRKLIRLGLDIDIESQTC 94
 DB 1 EELDGGCHLVTYODSGMTSKNPGTYPNHTVCEKITVTKRKLIRLGLDIDIESQTC 60

QY 95 ASDYLFTSSDDQGYPCGSMVTPEKLLNTSEVTFRESGSHISGRGFLTYASSDHPD 154
 DB 61 ASDYLFTSSDDQGYPCGSMVTPEKLLNTSEVTFRESGSHISGRGFLTYASSDHPD 120

QY 155 LITCLERASHLTKETYSKFCPCAGCRDVAAGDISGMNVDGRTSLCKAAIHAGIADBLG 214
 DB 121 LITCLERASHLTKETYSKFCPCAGCRDVAAGDISGMNVDGRTSLCKAAIHAGIADBLG 180

QY 215 GOIIVLORKGISREGILANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQOIRASSSMO 274
 DB 181 GOIIVLORKGISREGILANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQOIRASSSMO 240

QY 275 SVNBSGDVHNSPCQARLQDGPMSWAGSDSSNNHKPREMLEIDGKKKITGIRTTGSTQ 334
 DB 241 SVNBSGDVHNSPCQARLQDGPMSWAGSDSSNNHKPREMLEIDGKKKITGIRTTGSTQ 300

QY 335 SNFNYKSFPMNFKNNNSKKKTYKGIYNNEEKYPOGNSNFRDPVQNNFIPPIYARYRV 394
 DB 301 SNFNYKSFPMNFKNNNSKKKTYKGIYNNEEKYPOGNSNFRDPVQNNFIPPIYARYRV 360

QY 395 VPOTWHORIALKVELIGQITQGNDSLWKRKTSOSTSVSTKEDETRIPISPEETSGI 454
 DB 361 VPOTWHORIALKVELIGQITQGNDSLWKRKTSOSTSVSTKEDETRIPISPEETSGI 420

QY 455 N 455
 DB 421 N 421

RESULT 5
 US-10-003-132-4
 ; Sequence 4, Application US/10003132
 ; Publication No. US20020192750A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Gao, Zeren
 ; APPLICANT: Shoemaker, Kimberly E.
 ; TITLE OF INVENTION: NEUROPIILIN HOMOLOG ZCUB5
 ; FILE REFERENCE: 00-62
 ; CURRENT APPLICATION NUMBER: US/10/003.132
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/249,004
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-10-003-132-4

Query Match 51.0%; Score 1924.5; DB 9; Length 503;
 Best Local Similarity 55.8%; Pred. No. 5.9e-141;
 Matches 392; Conservative 34; Mismatches 73; Indels 203; Gaps 4;

QY 14 AAGRLALLAASAPRLQAELGIDGCHLVTYODSGTMTSKNPGTYPNHTVCEKITIT 73
 DB 5 AAGRLALLAASAPRLQAELGIDGCHLVTYODSGTMTSKNPGTYPNHTVCEKITIT 64

QY 74 VPKRRLIRLIGDDISQTCASDYLFTSSDDQGYPCGSMVTPEKLLNTSEVTVRFE 133
 DB 65 VPKRRLIRLIGDDISQTCASDYLFTSSDDQGYPCGSMVTPEKLLNTSEVTVRFE 124

QY 134 SGSHISGRGFLTYASSDHPDLITCLERASHLTKETYSKFCPCAGCRDVAAGDISGMNVDG 193

DB 125 SGSHISGRGFLTYASSDHPDLITCLERASHLTKETYSKFCPCAGCRDVAAGDISGMNVDG 184
 QY 194 RDTSLCKAAIHAGIADBLGGOISVLORKGISREGILANGVLSRDGSLDKRFLFTSN 253
 DB 185 RDTSLCKAAIHAGIADBLGGOISVLORKGISREGILANGVLSRDGSLDKRFLFTSN 241

QY 254 GCSNLSFEPDQOIRASSSMOVSNSGQVHNSPCQARLQDGPMSWAGSDSSNNHKPREW 313
 DB 242 ----- 241

QY 314 LEIDGKKKITGIRTTGOSTOSNFTYKSFVPMNFKNNNSKKKTYKGIYNNEEKYPOGNS 373
 DB 242 ----- 241

QY 374 NFRDPVQNNFIPPIYARYRVVQTMHORIALKVELIGQITQGNDSLWKRKTSOSTSVS 433
 DB 242 ----- 241

QY 434 TKKEDETRIPISPEETSGINTTVAIPLVLLVLPFAGMGIFPAFRKKKKKSPYGA 493
 DB 242 -----TTPGNITTTVAIPSVIFALLITGMGIFALCRKKKKKNPYVA 285

QY 494 ENAKTDCKKOIKYPPARHQSLEFTSYDNKEKMTOKIDLITSMDADYQOPLMIGTYTR 553
 DB 286 DAQTKGCKKQIKYPPARHQSLEFTSYDNKEKMTOKIDLITSMDADYQOPLMIGTYTR 345

QY 554 KGFTRPMDTAEAGYSTDAGHYDCPCQARREYALPLAPPEVATPIVERHVLRAH 613
 DB 346 KGFTRPMDTAEAGYSTDAGHYDCPCQARREYALPLAPPEVATPIVERHVLRAH 405

QY 614 TFSASGTRVPGPOGRKHSLSGGSFVAGVAGDGYORPHSAQAPDRGYDRPKAVSA 673
 DB 406 TFSOSGTRVPGPRPTHSHSSGSGFPATG-ATQVESYQRPASPKPVGGYDRPKAASSF 464

QY 674 LATESGHPDSQKPTTHGTSYSAPRDLPLNOTAFTALL 715
 DB 465 L-DSRDPASQSQMT-SGDDGYSAPRNGLAPLNOTAFTALL 503

RESULT 6
 US-09-823-038A-51
 ; Sequence 51, Application US/09823038A
 ; Patent No. US20020058335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Strachan, Iorna
 ; APPLICANT: Sleeman, Matthew
 ; APPLICANT: Abernethy, Nevin
 ; APPLICANT: Ornust, Rene
 ; APPLICANT: Kumbie, Anand
 ; APPLICANT: Murlison, Greg
 ; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
 ; FILE REFERENCE: 11000.1037c3
 ; CURRENT APPLICATION NUMBER: US/09/823.038A
 ; CURRENT FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 51
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: Mouse
 ; US-09-823-038A-51

Query Match 50.7%; Score 1912.5; DB 10; Length 503;
 Best Local Similarity 55.6%; Pred. No. 5e-140;
 Matches 390; Conservative 35; Mismatches 74; Indels 203; Gaps 4;

QY 14 AAGRLALLAASAPRLQAELGIDGCHLVTYODSGTMTSKNPGTYPNHTVCEKITIT 73
 DB 5 AAGRLALLAASAPRLQAELGIDGCHLVTYODSGTMTSKNPGTYPNHTVCEKITIT 64

QY 74 VPKRRLIRLIGDDISQTCASDYLFTSSDDQGYPCGSMVTPEKLLNTSEVTVRFE 133

Db 65 VPKGRLLRLGDLINISKTCADSYLLFSSATDQYGYCGSMVAPKELRLNNEVTYLFK 124
 QY 134 SGRHSRGRLLTYASSDHDLLITCLERASHYLTETSEKFCPCACRDVADISGNMVDY 193
 Db 125 SGRHSRGRLLTYASSDHDLLITCLERASHYLTETSEKFCPCACRDVADISGNMVDY 184
 QY 194 RDTSLCKAAIHAGIIDELEGQISVLRKRGISRYEGIILANGVLSRDSLSDRKFLTSN 253
 Db 185 RDTSLCKAAIHAGIIDELEGHINLQSKGISHYEGLANGVLSRHSLSERKFLF 241
 QY 254 GCSRSLSFEEDGOIRASSSQSVNESGDQYHWSPGARLQDQGPMSAGSSNNHKKREW 313
 Db 242 241
 QY 314 LEIDLGRKKITGIRTTGTSQSNFNYKSFVNFKNNSKMTYKGIYNNKEKVFQGS 373
 Db 242 241
 QY 374 NFRDPVQNNFIPPIVARYVRVPQTHQRIALAVELIGCOITQGNDSLVMKRTSQSTSVS 433
 Db 242 241
 QY 434 TKREDETIRPISEETSTGINTTVAIPVLVLVLFVAGMGIFAAFRKKKKGSPYGA 493
 Db 242 241
 QY 494 EAQKTDCKWKIKYPPFAHQAETIISTDNKEKMTOKLDLITSMADYQOPLMGTGTVR 553
 Db 286 DAQKTCGKWKIKYPPFAHQAETIISTDNKEKMTOKLDLITSMADYQOPLMGTGTVR 345
 QY 554 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHYALPLAPEPEXATPIVERHVLRAH 613
 Db 346 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHYALPLAPEPEXATPIVERHVLRAH 405
 QY 614 TFSAGSYRVGPQPGHKSLSGGFSPVAGVGAQDQYRPHSAQPADRGYDRPRAVSA 673
 Db 406 TFSAGSYRVGPQPGHKSLSGGFSPVAGVGAQDQYRPHSAQPADRGYDRPRAVSA 464
 QY 674 LATESGHPDSQKPTHTGTSDSYAPRDCLTPLNQTMTALL 715
 Db 465 L-DSRDPASQSQMT-SGGDGYSAAPRNGLAFLNQTMTALL 503
 RESULT 7
 US-10-003-132-6
 ; Sequence 6, Application US/10003132
 ; Publication No. US20020192750A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Gao, Zeren
 ; APPLICANT: Shoemaker, Kimberly E.
 ; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
 ; FILE REFERENCE: 00-62
 ; CURRENT APPLICATION NUMBER: US/10/003,132
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/249,004
 ; PRIOR FILING DATE: 2000-11-15
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FaastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 458
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-003-132-6

Query Match 44.8%; Score 1690; DB 9; Length 458;
 Best Local Similarity 50.8%; Pred. No. 7,8e-123;
 Matches 354; Conservative 31; Mismatches 69; Indels 248; Gaps 5;

QY 14 AAGRGTLALLAVSAPLQAEELGDCGHLVYODSGTWSKKNYPGTYPNHTVCEKTIIT 73
 Db 5 AGGPSVYALLFVACAPLRLQAEELGDCGHLVYODSGTWSKKNYPGTYPNHTVCEKTIIT 64

QY 74 VPKGRLLRLGDLINISKTCADSYLLFTSSDQYGYCGSMVAPKELRLNNEVTYLFK 133
 Db 65 VPKGRLLRLGDLINISKTCADSYLLFSSATDQYGYCGSMVAPKELRLNNEVTYLFK 99
 QY 134 SGRHSRGRLLTYASSDHDLLITCLERASHYLTETSEKFCPCACRDVADISGNMVDY 193
 Db 100 133
 QY 194 RDTSLCKAAIHAGIIDELEGQISVLRKRGISRYEGIILANGVLSRDSLSDRKFLTSN 253
 Db 140 RDTSLCKAAIHAGIIDELEGHINLQSKGISHYEGLANGVLSRHSLSERKFLF 196
 QY 254 GCSRSLSFEEDGOIRASSSQSVNESGDQYHWSPGARLQDQGPMSAGSSNNHKKREW 313
 Db 197 196
 QY 314 LEIDLGRKKITGIRTTGTSQSNFNYKSFVNFKNNSKMTYKGIYNNKEKVFQGS 373
 Db 197 196
 QY 374 NFRDPVQNNFIPPIVARYVRVPQTHQRIALAVELIGCOITQGNDSLVMKRTSQSTSVS 433
 Db 197 196
 QY 434 TKREDETIRPISEETSTGINTTVAIPVLVLVLFVAGMGIFAAFRKKKKGSPYGA 493
 Db 197 196
 QY 494 EAQKTDCKWKIKYPPFAHQAETIISTDNKEKMTOKLDLITSMADYQOPLMGTGTVR 553
 Db 241 DAQKTCGKWKIKYPPFAHQAETIISTDNKEKMTOKLDLITSMADYQOPLMGTGTVR 300
 QY 554 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHYALPLAPEPEXATPIVERHVLRAH 613
 Db 301 KGSFPRMDTDAEAGYSTAGHYDCPORAGREHYALPLAPEPEXATPIVERHVLRAH 360
 QY 614 TFSAGSYRVGPQPGHKSLSGGFSPVAGVGAQDQYRPHSAQPADRGYDRPRAVSA 673
 Db 361 TFSAGSYRVGPQPGHKSLSGGFSPVAGVGAQDQYRPHSAQPADRGYDRPRAVSA 419
 QY 674 LATESGHPDSQKPTHTGTSDSYAPRDCLTPLNQTMTALL 715
 Db 420 L-DSRDPASQSQMT-SGGDGYSAAPRNGLAFLNQTMTALL 458
 RESULT 8
 US-09-759-130B-78
 ; Sequence 78, Application US/09759130B
 ; Publication No. US2003002279A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Millennium Pharmaceuticals, Inc.
 ; APPLICANT: McCarthy, Sean A.
 ; APPLICANT: Fraser, Christopher C.
 ; APPLICANT: Sharp, John D.
 ; APPLICANT: Barnes, Thomas S.
 ; APPLICANT: Kirst, Susan J.
 ; APPLICANT: Mackay, Charles R.
 ; APPLICANT: Myers, Paul S.
 ; APPLICANT: Leiby, Kevin R.
 ; APPLICANT: Wrighton, Nicolas
 ; APPLICANT: Goodearl, Andrew
 ; APPLICANT: Holtzman, Douglas A.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
 ; PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
 ; FILE REFERENCE: MP100-5350N1M
 ; CURRENT APPLICATION NUMBER: US/09/759,130B
 ; CURRENT FILING DATE: 2002-09-16
 ; PRIOR APPLICATION NUMBER: US 09/479,249
 ; PRIOR FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: US 09/559,497
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: US 09/578,063

```
;; PRIOR FILING DATE: 2000-05-24
;; PRIOR APPLICATION NUMBER: US 09/333,159
;; PRIOR FILING DATE: 1999-06-14
;; PRIOR APPLICATION NUMBER: US 09/596,194
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 09/342,364
;; PRIOR FILING DATE: 1999-06-29
;; PRIOR APPLICATION NUMBER: US 09/608,452
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/393,996
;; PRIOR FILING DATE: 1999-09-10
;; PRIOR APPLICATION NUMBER: US 09/602,871
;; PRIOR FILING DATE: 2000-06-23
;; PRIOR APPLICATION NUMBER: US 09/420,707
;; NUMBER OF SEQ ID NOS: 460
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 78
;; LENGTH: 235
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-759-1308-78
```

```
Query Match      33.68; Score 1265; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 2,7e-90;
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 481 RKKKKSGPYSAAEQKTDCKMOKIYPPARHQAFTISYDNKEMTKLDLITSDMADY 540
Db 1 RKKKKSGPYSAAEQKTDCKMOKIYPPARHQAFTISYDNKEMTKLDLITSDMADY 60
OY 541 QOPLMIGTGYTRKSTFRPMDTDAEAGVSTDAGHYDCPQAGRHEYALPLAPEPEY 600
Db 61 QOPLMIGTGYTRKSTFRPMDTDAEAGVSTDAGHYDCPQAGRHEYALPLAPEPEY 120
OY 601 APTIYERVLAAHPTSAQSGYRVPGOPGKHRSLSGGSPVAGVAGDGDYORPHSAP 660
Db 121 APTIYERVLAAHPTSAQSGYRVPGOPGKHRSLSGGSPVAGVAGDGDYORPHSAP 180
OY 661 ADRGIDRKAVALATESGHPDSQKPTHPGTSYSAPRDLPLNOTAMTALL 715
Db 181 ADRGIDRKAVALATESGHPDSQKPTHPGTSYSAPRDLPLNOTAMTALL 235
```

```
RESULT 9
US-10-060-830-3
; Sequence 3, Application US/10060830
; Publication No. US20030032154A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
; FILE REFERENCE: PB0169
; CURRENT APPLICATION NUMBER: US/10/060,830
; PRIOR FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/325,062
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 1123
; SOFTWARE: Neomica Sequence Listing Engine
```

```
;; SEQ ID NO 3
;; LENGTH: 729
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-060-830-3
```

```
Query Match      27.68; Score 1042; DB 9; Length 729;
Best Local Similarity 36.4%; Pred. No. 2,7e-72;
Matches 263; Conservative 113; Mismatches 240; Indels 106; Gaps 23;
```

```
OY 20 LALLIYAVAPRLIAEELGDCGHLYTQDSGTMSSKRYPCPTYPHNTCENTITYPKGR 79
Db 5 LLLLVLLLLLEDGAGQDGGHVLCPESGTLISNVPOTVPSTVCEIEIRKMGSR 64
OY 80 LILRLGDDIE-SQTCADYLLFTS----SSDQYPCYG-SMTYPKELLMTSEYVFE 133
Db 65 VRIKRGDDIEDSDSCHNYLRIYNGIVSSTEIKCYGLGLQNMHSLESGNETLLFM 124
OY 134 SGSHISGRGFLITYASSHPDLITCLERASHYLTKEYSKFCPCAGCDVAGDISGMVDGY 193
Db 125 SGIHVSGRGLASYEVIDKODLITCLDPAISNLFEPESKCPACCLLPFAISGTIPHG 184
OY 194 RDTSLCKAAIHAGIADLDELGOISVLOKGISREGLANGVLSRDSLSDKRRLFTSN 253
Db 185 RDSPLCMAGVAGVSNLTGQISVLSKGIPTYESSLANNVTSVGHLSLTFERTS 244
OY 254 GCSRSLSFE---PDGQIRASSMSQSVNESGDQVWMSFGARLADQGPSMASGDSNNHK 309
Db 245 GCYGTGLNMGESVIADPQITASSVLEMTDHTGOENSMKRAKRPAPMAFATDE--- 301
OY 310 PREMLEIDLGKKKITGIRTTGSTOSNFNYKSFVNFKNNSKWKTYKGIIVNEEKVF 369
Db 302 -YQWQIOLNKEKKITGITITGSTVWEHNYVYSARILYSDGQKWTYRREPVEODKIF 360
OY 370 GNSNFRDPVONNFIPIYAVRVYVPOTVHORIALKYLIGCO-----ITQ----- 416
Db 361 GKNKDYHODVKNRNLPLIIRAFIRYNPTQWOOKIAMKMLGCOPIPKRPPKLTQPPPP 420
OY 417 --GNDSLVWRKTSQTSVSTRKEDETITRP-----PSEET-STGINITT----- 458
Db 421 RNSND---LNTTAPRIKAGRAKFTQPLQPSNSNEFPQOTEGTASPIRNTVPPN 476
OY 459 ----VAIPVLVLYVFGMGIFPAF-----RKKKKKS---PYCSAAEQKTDCKQ 503
Db 477 VTKDVALAVALVPLVWLTLILLLVLCAMHWRNKKRKEGTIDLPYV----DRAGMKG 532
OY 504 IKYFP-----ARHQAFTISYDNKEMTKLDLITSDMADYQOPLMIGTGYT 552
Db 533 MKQFLPAKAVDHEETPVKYSSE--VNHLSPREVT--TVLQADSAEYAOPLVGGIYCTL 587
OY 553 RKSTFRPMDTDAEAG-VSTDAGGHYDCPQAGRHEYALPLAPEPEYATPIV---ERH 608
Db 588 HQRSFTKFP--DEKRAVYADDP---VNSPQEVYHAAVEPLITGPPEYANPIIDMSGH 642
OY 609 VLRANTFSAGSGYRVPGOPGKHRSLSGGSPVAGVAGDGDYORPHSAPADRGYDRP 668
Db 643 PTVSGOPSTSTFKATGNP-----PPL--VGTYVTLISRTDSCSSAQAQYDTP 689
OY 669 KA 670
Db 690 KA 691
```

```
RESULT 10
US-10-060-830-1114
; Sequence 1114, Application US/10060830
; Publication No. US20030032154A1
; GENERAL INFORMATION:
; APPLICANT: Gu, Yizhong
; TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
; FILE REFERENCE: PB0169
; CURRENT APPLICATION NUMBER: US/10/060,830
```

```

; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/325,062
; PRIOR FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 1123
; SOFTWARE: Aemica Sequence Listing Engine
; SEQ ID NO 1114
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-830-1114

```

```

Query Match      23.2%; Score 873.5; DB 9; Length 653;
Best Local Similarity 35.1%; Pred. No. 2,6e-59;
Matches 224; Conservative 102; Mismatches 21; Indels 101; Gaps 21;

```

```

QY 99 LFTSSSDQGYPCG-SMYPKELLNTSEYVAFESGSHSGFLLTASSHPDLIT 157
DB 13 LLEDAAQGGKCGGLGNHNSIESKGNITLLFMGSHVSGFLASYSVIDKODLIT 72
QY 158 CLERASHYLTREXSKPCACRDVAGDISGMMVGYRDTSLCKAALHAGIADLGOI 217
DB 73 CLDTASFTLEPERSKYPRACCLPFAISGTFPHGIVSDSPLCAGVAGVSTLGGOI 132
QY 218 SVLQKGISREGLANGVLSRDGSLDKRFLFTSNCSLSLFE---PDGQIRASSSW 273
DB 133 SVYISKIPYESSLANNNVSVCHLSLFTFKTSCYGLMGESGVINDPOITASSVL 192
QY 274 QSVNESGDQVHMSGQARLDODGPRVNASGSSNNHAKREMLEIDLGKKITGRTGST 333
DB 193 EMTDHTGQEMSKPKRRLKPPGPMAPATDE---YOMQIDLNEKKITGRTGST 248
QY 334 QSNFNFYKSFVAMFNKNNKMKTYKGIYVNEEKVFQGNFNDPVOONFPIPIVATYR 393
DB 249 MVENHYVSAIRILYSDGQKTYRREPVEQDKIFQGNKDYHODVANNFLPIIAFIR 308
QY 394 VNPOTWRIALVELIGCO---ITQ---GNDSLWRTSOSTSVSTRKED 438
DB 309 VNPOTWRIALVELIGCOFIPGPRPKLTOPPPRNSND---LNTTAPRIANGRA 364
QY 439 EITTRP-----PSEET---STGINTT-----VALPVLVLVFAGMIFA 478
DB 365 PKFTQPLQPRSSNEFPQOTQTTASPDIRNTTYTPNTKDALAVALVLPVWLTLITL 424
QY 479 AF-----RKKKKKGS---PYGSABAQKTDCKWOLKYPF-----ARHOSAEF 516
DB 425 ILVCAHMRKRRKKTETETYLDPYV---DAGMKKMGKQOLPAKAVDHETTPVYSSE- 479
QY 517 TITDNEKERTOKLIDLTSMADYQOPLMTGTGTTRKGSFPRMDTAEAG-VSTDAG 575
DB 480 -VNLHLSREYV---TVQADSAEVAQPLVGIYVTLHQSTFRP--BEKREAGVADLP- 532
QY 576 GHVDCPQARHREYALPLAPPEPYATPIY--ERNHLRAHTPSAOGIYRPGQPHKH 632
DB 533 --TNSPQOEYVHAIAELPTTGPETATPIIMDSGHPHTTSVGQSTSTFRATGNP--- 586
QY 633 SLSSGFSVAGVAGDGDYQRPHSADPADRGYDRPKA 670
DB 587 -----PPL--VGTVTTLSTRDSCSSAAQADITPKA 615

```

```

RESULT 11
US-10-104-440-4
; Sequence 4; Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROFILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 4802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 931
; TYPE: PRT
; ORGANISM: human
US-10-104-440-4

```

```

Query Match      9.3%; Score 351.5; DB 12; Length 931;
Best Local Similarity 22.2%; Pred. No. 1.3e-18;
Matches 142; Conservative 99; Mismatches 255; Indels 143; Gaps 25;

```

```

QY 41 CGHLYTQDSGTMSKRYPTGTPMHTYCEKTIYPR-GKRLILRLG-DLIESQTCASDY 98
DB 28 CGGRLNSKDGAYTISPYDPDYPHONCEWIVYVAPENQIVLNFNPFIEIKDKYDR 87
QY 99 LTF-----TSSSDQGYPCGSMYVPEKELLNTSEYVAFESGSHSGFLLTY-----AS 149
DB 88 IEIHDGSESADLGRKCGNI-APPTIISGSMIYIFSTDYARQCGFSLRIEIRFTGS 146
QY 150 SD-----HPLITC-----LERASHYLTREXSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGPEKYPHNDCFTTILAKPKMELIQLFLFDEHPLQYCEG 206
QY 174 -CPAGCDVAGDIS--GNMVDGYRDTSLCKAALHAGIADLGOISVLRKGISREY 230
DB 207 DCKYDMDLMDGIPHYVGLIGKCYGTPTSELRSSTGLSLTFRTDMAVAK-----DG 259
QY 231 ILANGVLSRDGSLDKRFLFTSNCSLSLFE---PDGQIRASSMQSVNESGDQVHMS 286
DB 260 PSARYIYVHOPEPLENFO-----CNVPLMGESGRINEDISASTY-----SDGRWT 305
QY 287 PGARLADODGPRVNASGSSNNHAKREMLEIDLGKKITGRTGST--STGSNFFYKSF 344
DB 306 POGSRLHGDDNGWTPNLDNS---KEYLOYDLRLFTLMAIANGAISRTONGYVYKSY 361
QY 345 VAMFNKNNKMKTYKGIYVNEEKVFQGNFNDPVOONFPIPIVATYRVPOTWRIA 404
DB 362 KLEVSTGEDMAYRHRKNH--KVFOANNDAATEVVLKRLAPLITRVRIRPQTHWGIA 419
QY 405 LKVELICQITQGNDSLVWMTKS---OSTSVSTRKEDETTRP-----IP 446
DB 420 LRLELFGCRVTDAPCSNMLGMLSLINDSQISASSTOETVIMSSAANLVSRSGMFPRI 479
QY 447 SEETSTGINTTYVAIPVLVLFVAFAGM--IFA---AFRRKKKSGPYGSNAQKTD 499
DB 480 QAGPEEMLVQDICTPRTVAGVILIOGARGDSITAVBARAFVKKFYSYSLNGKD----- 534
QY 500 CMKQIKTFPARHOSA-EFTISYD-----NEKEWTKOKLIDLTSD 536
DB 535 -WEYIDPRTQOAPKLEFGNNHMDTPDIRRDPIDPAQYRVYPERWSPAGIGMLEVIGCD 593

```

QY 537 MADYQDPLMIGTGTTRKSGTFRPMQDADAEAGVSTDAG 575
 Db 594 WTD-SKPTVETLIGPTVKSEETTPYPTTEE-----ATRCG 627

RESULT 12

US-10-104-440-2
 ; Sequence 2, Application US/10104440
 ; Patent No. US20020132774A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KLAGSBRUN, Michael
 ; APPLICANT: SOKER, Shay
 ; APPLICANT: MIAO, Hua Quan
 ; TITLE OF INVENTION: ANTAGONISTS OF NEUROFILIN RECEPTOR FUNCTION AND USE
 ; FILE REFERENCE: 48802 C
 ; CURRENT APPLICATION NUMBER: US/10/104,440
 ; CURRENT FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: 09/580, 803
 ; PRIOR FILING DATE: 2000-05-30
 ; PRIOR APPLICATION NUMBER: 60/069, 155
 ; PRIOR FILING DATE: 1997-12-09
 ; PRIOR APPLICATION NUMBER: 60/069, 687
 ; PRIOR FILING DATE: 1997-12-29
 ; PRIOR APPLICATION NUMBER: 60/078, 541
 ; PRIOR FILING DATE: 1998-03-19
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 923
 ; TYPE: PRT
 ; ORGANISM: human
 ; US-10-104-440-2

Query Match 8.9%; Score 335; DB 12; Length 923;
 Best Local Similarity 24.9%; Pred. No. 2,4e-17;
 Matches 124; Conservative 76; Mismatches 176; Indels 122; Gaps 23;

QY 17 RGL-----LALLAVASAPRLQAEELGDCGHLVYQDSGTMTSKNYPGTYPNHTVCEK 70
 Db 3 RGLPLCALVALVLAAPGAFR-----NDGCGPTIKIESGYLSPGYRHSYPSKCEW 56
 QY 71 TTVVPGK-KRLILRLG-DLDIESQTCASDYL-LFTSSDO---YGPYCGSMVTPKELLN 124
 Db 57 LIDAPDYOQRIIMNFNFHLEDRDCKYDYVEVFDGENENGHRGKCGRI-APPVVS 115
 QY 125 TSEVTVFEESGSHISGKFLTYA-----SSDHDLLITC-- 158
 Db 116 GPFLEIFVSDYETHGAGFIRIRIEIRKRGPCSONTTTSGVTKSPGFPEKYPNSLECY 175
 QY 159 -----LERASHYLTKEYSKFCPCAG-CRDVAGDISGMVNDGYRDTSLCKRAIR 205
 Db 176 IVFAPKSEILIEFESDELPDSNP-PGGMPCRYDRLEI-----WQGFPP-----VGRH 223
 QY 206 AG-IIDLEGQI-----SVLQKRGISRYEISILANGVLSRDGSLSDKRFLE 250
 Db 224 IGRYCGKRTGRIRSSGILSMVFYTDISAIAKEGFSANYSV-SEDFK----- 274
 QY 251 TSNCGSRLSEFP---DQIRASSSMQSVNESGDQVHWSPGARLQDQGPASAGDSSN 306
 Db 275 -----CMBALMESGEIHSDDITASSQST-----NMSAERSRLNYPENGWTPBEDSY 322
 QY 307 NHRPREMLEIDLGEKKRTGIRTTG-STQSNFNFYKSVNMFKNNSKWKYKGIYNN 364
 Db 323 -----REMIQYDGLIRVYAVTGOAISKEFKKKYKTKKIDYSSNGEWMITIK--EGN 376
 QY 365 EEKVFQGSNFRDPVONNFIPIVARYVRVPQTMHQRALAKVELIGCOIT---OGNDS 420
 Db 377 KPVLFQGNINPTDYVAVFEPKPLITREVRIRKPAITWETGISMBREYVCKITIDPCSGMLG 436
 QY 421 LVWKTQSQTSVSTKKRD 438

Db 437 MWSGLISDSQITSSNOGD 454

RESULT 13

US-10-115-563-14
 ; Sequence 14, Application US/10115563
 ; Publication No. US20030008307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffard, John H
 ; APPLICANT: Greengard, Judith S
 ; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
 ; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
 ; AND COMPOSITIONS THEREOF
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: The Scripps Research Institute, Office of
 ; Patent Counsel
 ; STREET: 10666 NO. US20030008307A1th Torrey Pines Road, TPC 8
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/115,563
 ; FILING DATE: 02-APR-2002
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/410,488
 ; FILING DATE: 24-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fitting, Thomas
 ; REGISTRATION NUMBER: 34,163
 ; REFERENCE/DOCKET NUMBER: 449.0
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-554-2937
 ; TELEFAX: 619-554-6312
 ; INFORMATION FOR SEQ ID NO: 14:
 ; LENGTH: 2224 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
 ; US-10-115-563-14

Query Match 8.4%; Score 317; DB 9; Length 2224;
 Best Local Similarity 42.0%; Pred. No. 2,2e-15;
 Matches 71; Conservative 30; Mismatches 52; Indels 16; Gaps 5;

QY 253 NCGSRLSEFP---PDGQIRASS---SMQSVNESGDQVHWSFGARLQDQGPASAGDSS 305
 Db 2064 NCGSTPLGKENGKIKNKQITASSFKSW----GD--YMEFFARLAAQGRVNMQAKA 2116
 QY 306 NHRPREMLEIDLGEKKRTGIRTTG-STQSNFNFYKSVNMFKNNSKWKYKGIYNN 365
 Db 2117 NNKK--QMLEIDLKIKRTAITITOGCKSLSEMYKSTTTHTSOGVEMKPYRLKSSW 2174
 QY 366 EEKVFQGSNFRDPVONNFIPIVARYVRVPQTMHQRALAKVELIGCOI 414
 Db 2175 DKIFEGNRTKGVNKNFNPILISFIRIVPTWQSTILRLLEGCDI 2223

RESULT 14

US-10-095-718-4
 ; Sequence 4, Application US/10095718
 ; Patent No. US20020131956A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walsh, Christopher

APPLICANT: Chao, Hengjun
APPLICANT: Burslein, Haim
APPLICANT: Lynch, Carmel
APPLICANT: Stepan, Tony
APPLICANT: Munson, Keith
TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
FILE REFERENCE: 35052/204375
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/689,430
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/158,780
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1431
TYPE: PRT
ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match 8.1%; Score 306.5; DB 12; Length 1431;
Best Local Similarity 40.5%; Pred. No. 7.5e-15;
Matches 68; Conservative 26; Mismatches 61; Indels 13; Gaps 4;

QY 253 NCGRSLSE---PDGQIRASSSMQSVNEDQVHSPGOARLQDQPSWAGSDSSNNH 308
DB 1271 NCGRSLSEKRAISDAQTASSYLSMLAT-----WSPQARLHIGRTNMRQANN- 1324
QY 309 KPEWLEIDLGKKITGRTGTSQSNFNFYKSVFNNFKNNNSKMTYKGIYNNEKY 368
DB 1325 -KPEWLVDFKRTMKYTGITGVKSLISMYKEFLISSQDGHNTLF--LQNGKVKY 1381
QY 369 FQGNSEFRDPOVNNFPIVARYRVVPTWQRIALXELIGCQITQ 416
DB 1382 FQGNRDSSTPRNREPLVARYRLHPQSMHIALRLLEVLGCDTQ 1429

RESULT 15
US-10-190-593-2
Sequence 2, Application US/10190593
Publication No. US20030022221A1
GENERAL INFORMATION:
APPLICANT: LANGIT, Emanuel et al.
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
FILE REFERENCE: CL001246
CURRENT APPLICATION NUMBER: US/10/190,593
CURRENT FILING DATE: 2002-07-09
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 343
TYPE: PRT
ORGANISM: Human
US-10-190-593-2

Query Match 7.9%; Score 299; DB 9; Length 343;
Best Local Similarity 37.3%; Pred. No. 3.6e-15;
Matches 63; Conservative 33; Mismatches 55; Indels 18; Gaps 5;

QY 253 NCGRSLSE---PDGQIRASSSMQSVNEDQVHSPGOARLQDQPSWAGSD 303
DB 184 NCGRSLSEKRAISDAQTASSYLSMLAT-----WSPQARLHIGRTNMRQANN- 1324
QY 304 KPEWLEIDLGKKITGRTGTSQSNFNFYKSVFNNFKNNNSKMTYKGIYN 363
DB 239 KPEWLVDFKRTMKYTGITGVKSLISMYKEFLISSQDGHNTLF--LQNGKVKY 1381
QY 364 FQGNSEFRDPOVNNFPIVARYRVVPTWQRIALXELIGCQITQ 412

DB 295 GSSKIFPGMNDHSHKKNLEFPPIIARYVRLPVANHRIALRLLELIGC 343
Search completed: May 15, 2003, 13:21:32
Job time: 31.0681 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 : Search time 20.546 Seconds
(without alignments)
3345.478 Million cell updates/sec

Title: US-10-003-132-2

Perfect score: 3770
Sequence: 1 MYPGARGGAGALAAAGAGL.....YSAPRDCLPLNQMTALL 715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pIR_73:*
2: pIR1:*
3: pIR2:*
4: pIR3:*
5: pIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	331	8.8	2211 1	KFB05
2	330	8.8	927 1	J00948
3	321	8.5	2324 1	KRU05
4	318	8.4	2183 2	T42764
5	299	7.9	218 2	A47285
6	297.5	7.9	427 2	JC4915
7	285.5	7.6	216 2	A44258
8	285.5	7.6	2351 1	EZHU
9	276.5	7.3	463 1	A36479
10	276.5	7.3	2319 2	A47004
11	271.5	7.2	2133 2	T42763
12	268	7.1	401 2	S65138
13	268	7.1	427 2	S74211
14	259.5	6.9	409 2	T11743
15	217	5.8	845 2	JC5256
16	216.5	5.7	719 2	JC5256
17	192.5	5.1	1283 2	S51739
18	190	5.0	3133 2	T11799
19	186.5	4.9	686 1	S52093
20	174	4.6	1524 2	A59271
21	172.5	4.6	319 2	T15369
22	168	4.5	3623 2	T09456
23	165.5	4.4	986 1	B56788
24	165	4.4	730 1	BH01
25	165	4.4	823 1	A58788
26	163.5	4.3	449 2	A55362
27	162.5	4.3	819 2	I48859
28	160.5	4.3	694 2	JC5554
29	158	4.2	991 2	I49540

30	156	4.1	707 2	JC2218	procollagen C-endo
31	156	4.1	3623 2	T08618	intrinsec factor-B
32	155.5	4.1	855 2	S42621	protein-tyrosine k
33	153	4.1	402 2	JH0403	procollagen I C-pr
34	149.5	4.0	695 1	S05008	complement subcomp
35	142.5	3.8	1385 2	T14158	neurexin IV - mous
36	141	3.7	688 1	CIH05	complement subcomp
37	141	3.7	2083 2	T42721	CRP-ductin-alpha p
38	140.5	3.7	1381 2	T31083	paranodin - rat
39	137	3.6	705 1	CIH0RB	complement subcomp
40	136.5	3.6	699 1	I54763	ra-reactive factor
41	136	3.6	1019 2	A38738	coagulation factor
42	135	3.6	737 2	T15615	hypothetical prote
43	133.5	3.5	1594 2	T30549	hensin - rabbit
44	133	3.5	855 2	JC7731	membrane-bound arg
45	133	3.5	855 2	JC7775	membrane type-seri

ALIGNMENTS

RESULT 1
KFB05
coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C:Accession: A42580; A36497
R:Guinot, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A:Title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; M01D:92147638; PMID:1737753
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GU>
A:Cross-references: GB:M81440; NID:9163037; PIDN:AAA30512.1; PID:9163038
A:Note: Sequence extracted from NCBI backbone (NCBIN:80774, NCBI:80776)
R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A:Title: Identification and characterization of a phospholipid-binding site of bovine
A:Reference number: A36497; M01D:91072354; PMID:2254316
A:Accession: A36497
A:Molecule type: protein
A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584; 1673-1676, 'X', 1678-1679, 'X', 1681, 'X'
Biochemistry 33, 13109-13116, 1994
A:Title: Determination of the disulfide bridges in factor Va heavy chain.
A:Reference number: A55979; M01D:95034740; PMID:7947716
A:Contents: annotation
A:Note: 566-Cys and 617-Cys were shown to have free sulphydryls
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A:Pathway: blood coagulation
A:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:1-28/Domain: signal sequence #status predicted <S>
F:29-2211/Product: coagulation factor V #status predicted <M>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <A>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-695/Domain: A2 <DA2>
F:351-688/Domain: ferroxidase repeat homology <FO2>
F:696-1564/Domain: B <DOB>
F:1175-1437/Region: 9-residue repeats (Q-X-T/R-L-S-P-D-L-S)
F:1565-2211/Product: coagulation factor Va light chain #status predicted <A>
F:1565-2211/Product: coagulation factor Va light chain #status predicted <A>
F:1572-1892/Domain: A3 <DA3>
F:1572-1892/Domain: ferroxidase repeat homology <FO3>
F:1654-1152/Region: phospholipid binding #status predicted
F:1893-2051/Domain: C1 <DC1>
F:1893-2048/Domain: discoidin I amino-terminal homology <DN1>
F:2052-2211/Domain: C2 <DC2>
F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>
F:167-193, 248-329, 499-525/Disulfide bonds: #status experimental

A>Note: parts of this sequence were determined by protein sequencing
 R.Keller, F.G.; Ortel, T.L.; Quinlan-Allen, M.A.; Kane, W.H.
 Biochemistry 34, 4118-4124, 1995
 A>Title: Thrombin-catalyzed activation of recombinant human factor V.
 A:Reference number: A56139; MID:95210278; PMID:7696276
 A:Contents: annotation: thrombin cleavage sites
 C:Comment: factor V is activated by thrombin and partially by coagulation factor Xa.
 C:Genetics:
 A:Gene: GDB:F5
 A:Cross-references: GDB:119896; OMIM:227400
 A:Map position: 1q23-1q23
 A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prothrombinase
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:29-2224/Product: coagulation factor V #status predicted <MAT>
 F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAM>
 F:29-345/Domain: A1 <DA1>
 F:33-329/Domain: ferroxidase repeat homology <FO1>
 F:346-691/Domain: A2 <DA2>
 F:351-684/Domain: ferroxidase repeat homology <FO2>
 F:692-1573/Domain: B <DOB>
 F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
 F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
 F:1574-1905/Domain: A3 <DA3>
 F:1581-1905/Domain: ferroxidase repeat homology <FO3>
 F:1667-1765/Region: phospholipid binding #status predicted
 F:1906-2064/Domain: C1 <DC1>
 F:1906-2064/Domain: discoidin I amino-terminal homology <DNI>
 F:2065-2224/Domain: C2 <DC2>
 F:2065-2221/Domain: discoidin I amino-terminal homology <DNI>
 F:51-55239, 297, 460, 468, 554, 741, 752, 760, 776, 782, 821, 938, 977, 1074, 1083, 1103, 1106, 1479, 149
 F:167-193, 248-329, 500-526, 603-684, 1725-1751, 1907-2061, 2066-2221/Disulfide bonds: #status
 F:334-333/Cleavage site: Arg-Asn (protein C) #status predicted
 F:363-693, 1546/Binding site: sulfatase (Tyr) (covalent) #status predicted
 F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
 F:382-1338/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
 F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
 F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match. 8.5%; Score 321; DB 1; Length 2224;
 Best Local Similarity 42.6%; Pred. No. 3, 9e-14;
 Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;

OY 253 NGCSRSLSFE---PDGQIRASS---SWOSVNSGDVHNSPGQARLQDGPMSAGDSS 305
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 2064 NCSTPLGLEDGRIRQDQIRASSFKSMW----GD--YWEPLARLNAGRVANAWAKA 2116
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 OY 306 NNKHPREMLEIDLEKKKRTGIRTTGSGTOSNFNFYKSFVNFKNKNNKSKRTKGYVNE 365
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 2117 NNKK--QWLEIDLLKTKTKTAITQCKSLSEMYKSTHITSEGVKPYRLKSSMV 2174
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 OY 366 EKYFGNSNFRDPVONNFPIPIVARYRVVQPTWGORIALKVELICQI 414
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 2175 DKIFEGNSNTKGHKNFNFPIISRTIRIYPIKTNOSIALRLLEFGDI 2223
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 4
 T42764
 coagulation factor V - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T42764
 R:Yang, T.L.; Cui, J.; Rehmanulla, A.; Yang, A.; Mousaelli, M.; Kaufman, R.D.; Ginsburg,
 Blood 91, 4553-4599, 1998
 A>Title: The structure and function of murine factor V and its inactivation by protein C
 A:Reference number: Z22270; MID:98282202; PMID:9616155
 A:Accession: T42764

A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2183 <YAN>
 A:Cross-references: EMBL:U52925; NID:93219690; PID:93219691; PIDN:AAC99553.1
 C:Function:
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
 F:1541-1864/Domain: ferroxidase repeat homology <FOX>
 F:1541-1864/Domain: ferroxidase repeat homology <FOX>

Query Match. 8.4%; Score 318; DB 2; Length 2183;
 Best Local Similarity 40.8%; Pred. No. 6, 2e-14;
 Matches 69; Conservative 31; Mismatches 53; Indels 16; Gaps 5;

OY 253 NGCSRSLSFE---PDGQIRASS---SWOSVNSGDVHNSPGQARLQDGPMSAGDSS 305
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 2023 NCSTPLGLEDGRIRQDQIRASSFKSMW----GD--YWEPLARLNAGRVANAWAKA 2075
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 OY 306 NNKHPREMLEIDLEKKKRTGIRTTGSGTOSNFNFYKSFVNFKNKNNKSKRTKGYVNE 365
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 2076 NNKK--QWLEIDLLKTKTKTAITQCKSLSEMYKSTHITSEGVKPYRLKSSMV 2133
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 OY 366 EKYFGNSNFRDPVONNFPIPIVARYRVVQPTWGORIALKVELICQI 414
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 2134 DKIFEGNSNTKGHKNFNFPIISRTIRIYPIKTNOSIALRLLEFGDI 2182
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 5
 A47285
 milk fat globule protein - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
 C:Accession: A47285
 R:Carocca, D.; Peterson, J.A.; Ureia, R.; Kuniyoshi, J.; Bistrain, A.M.; Certant, R.L
 Cancer Res. 51, 4994-4998, 1991
 A>Title: A M-r 46,000 human milk fat globule protein that is highly expressed in huma
 A:Reference number: A47285; MID:91371351; PMID:1909932
 A:Accession: A47285
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <YAN>
 A:Cross-references: GB:S56151; NID:9235396; PIDN:AAB19771.1; PID:9235397
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
 F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DNI>
 F:60-218/Domain: discoidin I amino-terminal homology <DNI>

Query Match. 7.9%; Score 299; DB 2; Length 218;
 Best Local Similarity 37.3%; Pred. No. 4, 5e-14;
 Matches 63; Conservative 33; Mismatches 55; Indels 18; Gaps 5;

OY 253 NGCSRSLSFE---PDGQIRASSSWOSVNSGDVHNSPGQARLQDGPMSAGDSS 303
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 59 NGCANPLGLKNNKNSIPDQIRASSSYKWC---LHFSNPSYARLDQGNFANAWAGS 113
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 OY 304 SSNNHPRMLEIDLEKKKRTGIRTTGSGTOSNFNFYKSFVNFKNKNNKSKRTKGYVNE 363
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 114 YGND---QWLEIDLLKTKTKTAITQCKSLSEMYKSTHITSEGVKPYRLKSSMV 169
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 OY 364 NEKYFGNSNFRDPVONNFPIPIVARYRVVQPTWGORIALKVELICQI 412
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 170 GSSKIFPGNMNDNSHKKNLEFETILARYVRLVAVHNRIALRLLELGC 218
 :||:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 6
 JCA915
 ags protein precursor - rat
 N:Alternate names: O-acetyl-Gd3 ganglioside
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
 C:Accession: JCA915
 R:ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
 Biochem. Biophys. Res. Commun. 225, 932-938, 1996

A:Reference number: A43986; M01D:89340500; PMID:2503509
 A:Accession: A43986
 A:Molecule type: protein
 A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <FAY>
 R:Evsey, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.; J. Biol. Chem. 266, 740-746, 1991
 A:Title: Sulfation of Tyr(1860) of human blood coagulation factor VIII is essential for A:Contents: annotation; sulfation
 R:Gitschler, J.; Wood, W.I.; Goralika, T.M.; Wlton, K.L.; Chen, E.Y.; Eaton, D.H.; Vekhar, Nature 332, 326-330, 1984
 A:Title: Characterization of the human factor VIII gene.
 A:Reference number: A56196; M01D:85061547; PMID:6438525
 A:Contents: annotation; Introns
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M. Protein Sci. 4, 740-746, 1995
 A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of A:Reference number: A56216; M01D:95338127; PMID:7613471
 A:Contents: annotation; disulfide bonds
 A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
 R:Fjalke, M.; Hedling, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M. Eur. J. Biochem. 234, 773-779, 1995
 A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A:Reference number: S63527; M01D:96163459; PMID:8575434
 A:Accession: S63527
 A:Molecule type: protein
 A:Residues: 733-752;753-759 <KJA>
 R:Rind, P.; Larsson, K.; Sjöström, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg Eur. J. Biochem. 237, 19-27, 1995
 A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction A:Reference number: S66445; M01D:96048024; PMID:7556150
 A:Accession: S66445
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1668-1685 <LIN>
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure to C:Genetics: GDB:F8C
 A:Gene: GDB:F8C
 A:Cross-references: GDB:119124; OMIM:306700
 A:Map position: Xq28-Xq28
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro A:Pathway: Blood coagulation
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pld F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-335/Product: coagulation factor VIII #status experimental <MAT>
 F:20-740/Product: coagulation factor VIIIa heavy chain #status experimental <ACH>
 F:20-356/Domain: A1 <DA1>
 F:23-348/Domain: ferroxidase repeat homology <FO1>
 F:392-759/Domain: A2 <DA2>
 F:402-730/Domain: ferroxidase repeat homology <FO2>
 F:1668-1667/Domain: B <DB0>
 F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACLI>
 F:1709-2038/Domain: A3 <DA3>
 F:1716-2038/Domain: ferroxidase repeat homology <FO3>
 F:2039-2191/Domain: C1 <DC1>
 F:2039-2188/Domain: discoidin I amino-terminal homology <DN1>
 F:2199-2351/Domain: C2 <DC2>
 F:2199-2345/Domain: discoidin I amino-terminal homology <DN2>
 F:60, 558, 601, 776, 803, 847, 919, 962, 982, 1020, 1074, 1085, 1204, 1274, 1278, 1301, 1319, 1403, 1 F:172-198, 267-348, 547-573, 649-730, 1851-1877, 1918-1922, 2040-2188/Disulfide bonds: #status F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F:365, 737, 738, 742, 1683, 1699/Binding site: sulfate (Tyr) (covalent) #status experimental
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:414, 420/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1667-1668/Cleavage site: Arg-Glu (undefined proteolysis) #status experimental
 F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F:2199-2345/Disulfide bonds: #status predicted

Query Match 7.6%; Score 285.5; DB 1; Length 2351;
 Best Local Similarity 37.6%; Pred. No. 1.5e-11;
 Matches 62; Conservative 32; Mismatches 58; Indels 13; Gaps 4;
 253 NCGSRSLSEF-----PQOIRASSSMOSVNSGQVHSPQOALQOQSPWAGSDSSNNH 308
 Db 2191 NCSMPGLGMSKAIISDQITASSFTWFAF-----WSPKARHLQGRSNARPOVNN- 2244
 QY 309 KPREMLEIDGKKKIGITGTGTSNFMFVYKSPFANFNKNNKSKYTKGIYNNDEKY 368
 Db 2245 -PREMLQVDQKMKTKYGVTTGGVSKLITSMYKEFLISSQGHQWTLF--FQNGKVKY 2301
 QY 369 FQGNSEFDPVQNNFIPPIVARYRVVQPTWQRIALKEVLICQ 413
 Db 2302 FQGNSEFDPVQNNFIPPIVARYRVVQPTWQRIALKEVLICQ 2346
 RESULT 9
 A36479 milk fat globule membrane protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A36479
 R:Stubs, J.D.; Lekutis, C.; Singer, K.L.; But, A.; Yuzuki, D.; Strinivasan, U.; Parry Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
 A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the A:Reference number: A36479; M01D:91046008; PMID:2122462
 A:Accession: A36479
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-463 <STU>
 A:Cross-references: GB:M38337; NID:q199142; PIDN:AAA9534.1; PID:q199143
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom C:Keywords: membrane protein
 F:28-60/Domain: EGF homology <EG1>
 F:68-107/Domain: EGF homology <EG2>
 F:147-303/Domain: discoidin I amino-terminal homology <DN1>
 F:307-463/Domain: discoidin I amino-terminal homology <DN2>
 Query Match 7.3%; Score 276.5; DB 1; Length 463;
 Best Local Similarity 23.4%; Pred. No. 5.8e-12;
 Matches 112; Conservative 59; Mismatches 118; Indels 189; Gaps 21;
 94 CASDYLLFTSSSDQYGYCGSMTPPKELLNTSEVTVRFSGSHISGR-----GP 143
 Db 16 CASG--LFASGD---FCDS-----SLGLN-----GGTCLGQDNDIYCLDEGF 55
 QY 144 LLYTAS-----SDHDLITCLERASHYLTKEYSKCPACGRVADISGM- 189
 Db 56 TGLVCMETTERGPCSPNCPYNDACLVY--LDTORGDIPTFYICCPVGYSGICHCETETNY 114
 QY 190 -VDG-YRDTSLCKAAI-----HAGTIYD----- 211
 Db 115 NLGGEYFTTAVPTAAPTAPPPDLSNNLASKCSQGLGEGGALDSQISASTYVMGEW 174
 QY 212 ---ELGQISLVLRKGI-----SRYEGIIANGVLSRDG----- 241
 Db 175 GLGRMGPELRLRYRTGVNMAHSAVNSDKPMIYVNLKRKRVSGVMTQGA--SRAGRAEYL 233
 QY 242 -----SLSDKRPFLFTSN-----GCSR- 257
 Db 234 KTFKVAVYSIDGRKFEFIDSGDGKEFLGNDNNSIKVNNFNPTLEAQYIRLYPVSCHRG 293
 QY 258 -SLSEF-----PDGIRASSSMOSVNSGQVHSPQOALQOQSPWAGSDSSNNH 308
 Db 294 CTRFELGCEHLGCEPLGKLNNTIPDSOMSSSTKTNLRA--FGWTPHLGRIDNQG 351
 QY 297 P--SMASGDSNNHKKPREMLEIDGKKKIGITGTGTSNFMFVYKSPFANFNKNNKSKYTKGIYNNDEKY 354
 Db 352 KINAM-----TAQNSAKEMQLQVDLGRQVYTGITGGARDGFHIGVYVESKVAHSDQVQ 407
 QY 355 WKYIKGVINNEEVFQGNSEFDPVQNNFIPPIVARYRVVQPTWQRIALKEVLICQ 412

A:Accession: S78114
 A:Molecule type: protein
 A:Residues: 19-55;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
 R:Kim, D.H.; Kanno, C.; Mizokami, Y.
 Biochim. Biophys. Acta 1132, 203-211, 1992
 A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
 A:Reference number: S23926; M0ID:92353107; PMID:1643094
 A:Accession: S24181
 A:Molecule type: protein
 A:Residues: 383-394 <KIM>
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
 C:Keywords: blocked amino-end; disulfide bond; glycoprotein; milk
 F:1-16/Domain: signal sequence #status predicted <Sig>
 F:19-427/Product: PAS-6/7 protein #status experimental <Mat>
 F:24-58/Domain: EGF homology <EG1>
 F:66-105/Domain: EGF homology <EG2>
 F:108-265/Domain: discoidin I amino-terminal homology <DN1>
 F:269-427/Domain: discoidin I amino-terminal homology <DN2>
 F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted
 F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
 F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
 F:59;227/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:109-265;252-256;270-427/Disulfide bonds: #status experimental

Query Match 7.1%; Score 268; DB 2; Length 427;
 Best Local Similarity 36.7%; Pred. No. 2.1e-11;
 Matches 61; Conservative 30; Mismatches 63; Indels 12; Gaps 4;

OY 253 NGCSRLSEF---PDGIRASSQSVNESGQVHMSPGQALDQGP--SMASGSSN 306
 DB 268 NGCTPELEIDLEKRRKRTGRTGTSQSNFNYKSFVNFKNNSKMTYGIYVNEE 325
 OY 307 NHRPREMELEIDLEKRRKRTGRTGTSQSNFNYKSFVNFKNNSKMTYGIYVNEE 366
 DB 326 S---EMLQIDLGSGKRVYGIITQGARDFGHQYVAAYRAVAGDDGVTWEYDGPASGS 381
 OY 367 KYVQGSNFRDPVQNNFPIPIVARYRVVPTQTHQRIALVELIGC 412
 DB 382 KIFPGNNDNNSHKKNIFETPFQARFRIQVAVAHNRITLVELLGC 427

RESULT 14

T11743
 P47 protein - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T11743
 R:Enslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
 Biol. Reprod. 58, 1057-1064, 1998
 A:Title: Molecular cloning and characterization of P47, a novel boar sperm-associated z
 A:Reference number: Z17325; M0ID:98206817; PMID:9546740
 A:Accession: T11743
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-409 <ENS>
 A:Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA723/9.1; PID:g2652928
 A:Experimental source: testis
 C:Function:
 A:Description: may be involved in membrane remodeling and/or function as a zona pellucida
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:6-40/Domain: EGF homology <EGF>

Query Match 6.9%; Score 259.5; DB 2; Length 409;
 Best Local Similarity 25.0%; Pred. No. 7.9e-11;
 Matches 87; Conservative 46; Mismatches 102; Indels 113; Gaps 11;

OY 168 TYSKRCPCAGCRVADISNMVDGTRDTSL-----LCKAIIHGIIT-----ADELGC--- 215
 DB 72 TETICRCPHGYTGIIHCEIICNAPLGKMETGAIDFOISASSMHLGFMGLQRMABELARLHR 131
 OY 216 -----QISVLQKGISRYEGIIANGVLSRDGSL----- 243
 DB 132 AGIVNMTASNYDRNMIQVNLRR---KAVTVVIOGA-SRAGSAEYKTFKVAISTDG 187

OY 244 -----SDKREL-----FTSMGC 255
 DB 188 RKQFIQGAESDRIKIFMGLDNLGKLVNLFEPVLEQVYRVLPICHRCTLRFELLC 247
 OY 256 SRLSEFPDGS-----QIRASSQSVNESGQVHMSPGQALDQGP--SMASGDS 304
 DB 248 ELSCAEPLEIDLEKRRKRTGRTGTSQSNFNYKSFVNFKNNSKMTYGIYVNEE 305
 OY 305 NHRPREMELEIDLEKRRKRTGRTGTSQSNFNYKSFVNFKNNSKMTYGIYVNEE 364
 DB 306 SAS---EMLQIDLGSGKRVYGIITQGARDFGHQYVAAYRAVAGDDGVTWEYDGPASGS 361
 OY 365 EEFVQGSNFRDPVQNNFPIPIVARYRVVPTQTHQRIALVELIGC 412
 DB 362 EGKIFPGNNDNNSHKKNIFETPFQARFRIQVAVAHNRITLVELLGC 409

RESULT 15

JCS256
 adipocyte transcription factor, AEBP1 - human
 C:Species: Homo sapiens (man)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: JCS256
 R:Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo,
 Biochem. Biophys. Res. Commun. 228, 411-414, 1996
 A:Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expr
 A:Reference number: JCS256; M0ID:97079196; PMID:8920928
 A:Accession: JCS256
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-845 <OHN>
 A:Cross-references: DBJ:D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Query Match 5.8%; Score 217; DB 2; Length 845;

Best Local Similarity 25.7%; Pred. No. 2.6e-07;
 Matches 55; Conservative 40; Mismatches 97; Indels 22; Gaps 5;

OY 264 DGOIRASSSWQ--SVNESGQVHMSPGQARLDQGPVSSAGSSNNHNPRLMELEIDLEK 322
 DB 84 DNGIRASSMLRHGLGARGRLNMQGTATEDDYDGMACADDART-----QWIEVDTRRT 139
 OY 323 KITGIRTGSTQSNFNYKSFVNFKNNSKMTYGIYVNEEKFQGSNFRDPVQNN 382
 DB 140 RFTGVITQGRDSSIHDFVTTFVFGSNDGQVWMTN--GYDEMTFGHVDKDTPLSE 197
 OY 383 FIPPIVARYRVVPTQTHQRIALVELIGQITQ-----GNDSLVNR-----KTS 427
 DB 198 LPEPVARFRIYPLTWNGSLCKRLEVLGCSVAPVSYAQNENVAITDDIDFRHSTYKDM 257
 OY 428 QSTSVSTKKEDETITRPIPEESTSGINTVTAI 461
 DB 258 RQLMKVNECEPITITRIYSLGKSSRLKITAMEI 291

Search completed: May 15, 2003, 13:16:51
 Job time : 36.546 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:36 ; Search time 16.4368 Seconds
(without alignments)
1279.897 Million cell updates/sec

Title: US-10-003-132-2
Perfect score: 3770
Sequence: 1 MPPGARGGALRAAGRGILL.....YSAPRDLPLPLNOTATALL 715

Scoring table: BIOSUM62
Gapco 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCOTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	360.5	9.6	925	4	US-09-116-473-2 Sequence 2, Appl1
2	354.5	9.4	901	3	US-08-936-135-22 Sequence 22, Appl1
3	354.5	9.4	906	3	US-08-936-135-24 Sequence 24, Appl1
4	354.5	9.4	909	3	US-08-936-135-8 Sequence 8, Appl1
5	354.5	9.4	909	3	US-08-936-135-10 Sequence 10, Appl1
6	354.5	9.4	914	3	US-08-936-135-12 Sequence 12, Appl1
7	354.5	9.4	926	3	US-08-936-135-14 Sequence 14, Appl1
8	354.5	9.4	931	3	US-08-936-135-16 Sequence 16, Appl1
9	351.5	9.3	909	3	US-08-936-135-18 Sequence 18, Appl1
10	351.5	9.3	926	3	US-08-936-135-20 Sequence 20, Appl1
11	338.5	9.0	923	4	US-08-936-135-6 Sequence 6, Appl1
12	331	8.8	922	4	US-09-116-473-4 Sequence 4, Appl1
13	331	8.5	218	1	US-07-607-538C-4 Sequence 4, Appl1
14	321	8.5	218	2	US-08-162-402B-4 Sequence 4, Appl1
15	318	8.4	2183	3	US-08-746-111-5 Sequence 5, Appl1
16	311	8.2	160	2	US-08-162-402B-14 Sequence 14, Appl1
17	306.5	8.1	2343	4	US-09-324-867-2 Sequence 2, Appl1
18	299	7.9	217	1	US-07-607-538C-3 Sequence 3, Appl1
19	299	7.9	217	1	US-08-162-402B-3 Sequence 3, Appl1
20	289	7.9	218	1	US-07-607-538C-2 Sequence 2, Appl1
21	289	7.9	218	2	US-08-162-402B-2 Sequence 2, Appl1
22	299	7.9	387	2	US-08-162-402B-6 Sequence 6, Appl1
23	299	7.9	465	2	US-08-162-402B-8 Sequence 8, Appl1
24	293	7.8	159	2	US-08-162-402B-12 Patent No. 5422260
25	285.5	7.7	2351	6	US-09-209-916-1 Sequence 1, Appl1
26	285.5	7.6	1438	4	US-08-683-839B-3 Sequence 3, Appl1
27	285.5	7.6	1471	1	US-08-683-839B-3 Sequence 3, Appl1

28	285.5	7.6	1661	2	US-08-882-083-2 Sequence 2, Appl1
29	285.5	7.6	1661	2	US-08-558-107-2 Sequence 2, Appl1
30	285.5	7.6	1661	4	US-09-243-539-2 Sequence 2, Appl1
31	285.5	7.6	2332	1	US-07-864-004B-4 Sequence 4, Appl1
32	285.5	7.6	2332	1	US-08-251-937A-4 Sequence 4, Appl1
33	285.5	7.6	2332	1	US-08-212-133A-2 Sequence 2, Appl1
34	285.5	7.6	2332	1	US-08-276-594A-2 Sequence 2, Appl1
35	285.5	7.6	2332	1	US-08-474-503-2 Sequence 2, Appl1
36	285.5	7.6	2332	4	US-08-670-707A-2 Sequence 2, Appl1
37	285.5	7.6	2332	4	US-09-324-867-3 Sequence 3, Appl1
38	285.5	7.6	2332	4	US-09-315-179-2 Sequence 2, Appl1
39	285.5	7.6	2332	4	US-09-523-656-2 Sequence 2, Appl1
40	285.5	7.6	2332	5	PCR-US93-03275-4 Sequence 4, Appl1
41	285.5	7.6	2332	5	PCR-US94-13200-2 Sequence 2, Appl1
42	285.5	7.6	2351	1	US-08-121-202-2 Sequence 2, Appl1
43	285.5	7.6	2351	1	US-08-366-851A-2 Sequence 2, Appl1
44	285.5	7.6	2351	6	Patent No. 5171844
45	285.5	7.6	2351	6	Patent No. 5171844

ALIGNMENTS

RESULT 1
US-09-116-473-2
Sequence 2, Application US/09116473
Patent No. 6428965
GENERAL INFORMATION:
APPLICANT: Kolodkin, Alex
TITLE OF INVENTION: SEMAPHORIN RECEPTOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff
STREET: 1001 G Street, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/116,473
FILING DATE: 17-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/052,762
FILING DATE: 17-JUL-1998
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32141
REFERENCE/DOCKET NUMBER: 01107,74973
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 925 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-116-473-2
Query Match 9.6%; Score 360.5; DB 4; Length 925;
Best local Similarity 22.5%; Pred. No. 7.4e-23;
Matches 144; Conservative 99; Mismatches 246; Indels 151; Gaps 25;
QY 41 CGHLVYQDSGWTMSKNYPETYNHVCETIVPK-GKRLILRLG-DLDIESQTCASDY 98
|| : :|:|:| |||:| :| : :|:| :|:|

STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: UC97-288-2
REFERENCE/DOCKET NUMBER: 36,627
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-24

Query Match 9.4%; Score 354.5; DB 3; Length 906;
Best Local Similarity 21.5%; Pred. No. 2.4e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

QY 41 CGHLYTQDSGTMSTKNTPTGYPHNYCEKTIYTPK-GKRLILRLG-DLDIESQTCASDY 98
DB 28 CGGRPNKSDAGYITSPYPODYPSHONCEWIVVAPENOKIVLNFNPFLEKDKYDF 87
QY 99 LLF-----TSSSDQGYPCGSMWVPEKLLNTSEVTVPEESGSHSGRGLTTY-----AS 149
DB 88 IEIRDGSESDLLGKRCGN-APPTIISGSVYIYFTSDYANOGAGFSILRYEIFRTGS 146
QY 150 SD-----HPLDITC-----LERASHYLKTEYSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGFEKYPHNDCFTTILAKPRMEIILQFLFDEHDPLOYGEG 206
QY 174 -CPAGCDVAGDIS--GNMDGYDPTSLCKAAIHAGIADLGQISVLOKRSIKRY-- 228
DB 207 DCKYDMLDIMGIPHVGPLGKYCGTTPSKRLRSSTGILSTFTDMAVAVADGFSARYYL 266
QY 229 -----EGLANGVLSRDGSLSDKRFLLFTSNCGSRLSPFPGQIRAS 270
DB 267 IHQEPENFOCNVPLGMEGRINAEQISASTFSDGR----- 303
QY 271 SSWOSVNESGDQVHMSFGQARLADQGPSWASGDSSNNHKPREMLEIDLGEKKKITGIRTT 330
DB 304 -----WTPQGRSLHGDNGMTPNLDSN-----KEYLQVDFRFLMTALATQ 345
QY 331 G--STOSNFPYVYSFPMNKRNNNSKKTKYKGIYNNNEKYPOGNSNRDPQNNFPIPIV 388
DB 346 GAISREFOKGYVYSYLEVSTNGEDMAYVRHGKNH--KIFQANDATTEVVLNLHMFLL 403
QY 389 ARYRVVPTQTHORIALKVELIGCOITGNDLSVWRKTS---OSTSVSTKKEDEITRPI 445
DB 404 TRFIRINPQWHLGIALRLLEFGCRVTDAPCSNNLGMLSGLIADTQISASTREIYLSPS 463
QY 446 -----PSEETSTGINTTVAIPLVLLVAVFAGMGIFA-----AFRK 482
DB 464 AARLYSRSGWFPNPOAGGEEW-LQYDLCTPKTVKGVIIQARGGSDITAVBARAFVR 522
QY 483 KKKKGSYGSLEAKTKOCCKQIKYFPAHQSA-EFTISYD----- 521
DB 523 KFKVSYSLNGD-----WEYIDPRTQOQTKLFGNNHMYDTPDIRRDPVPAQYRVRYPE 576

QY 522 --NEKEWTKLDLITSDMADYQOPL-MIGETVTRKSGTFRPMDTDAEAG 569
DB 577 RMSPAGIGMRLEVLCGMDTBSKPTVEVLGPIVVKSEFTTTPRMDMDATECG 627

RESULT 4
US-08-936-135-8
Sequence 8, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: UC97-288-2
REFERENCE/DOCKET NUMBER: 36,627
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-8

Query Match 9.4%; Score 354.5; DB 3; Length 909;
Best Local Similarity 21.5%; Pred. No. 2.4e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

QY 41 CGHLYTQDSGTMSTKNTPTGYPHNYCEKTIYTPK-GKRLILRLG-DLDIESQTCASDY 98
DB 28 CGGRPNKSDAGYITSPYPODYPSHONCEWIVVAPENOKIVLNFNPFLEKDKYDF 87
QY 99 LLF-----TSSSDQGYPCGSMWVPEKLLNTSEVTVPEESGSHSGRGLTTY-----AS 149
DB 88 IEIRDGSESDLLGKRCGN-APPTIISGSVYIYFTSDYANOGAGFSILRYEIFRTGS 146
QY 150 SD-----HPLDITC-----LERASHYLKTEYSKF----- 173
DB 147 EDCSKNFTSPNGTIESPGFEKYPHNDCFTTILAKPRMEIILQFLFDEHDPLOYGEG 206
QY 174 -CPAGCDVAGDIS--GNMDGYDPTSLCKAAIHAGIADLGQISVLOKRSIKRY-- 228
DB 207 DCKYDMLDIMGIPHVGPLGKYCGTTPSKRLRSSTGILSTFTDMAVAVADGFSARYYL 266
QY 229 -----EGLANGVLSRDGSLSDKRFLLFTSNCGSRLSPFPGQIRAS 270
DB 267 IHQEPENFOCNVPLGMEGRINAEQISASTFSDGR----- 303
QY 271 SSWOSVNESGDQVHMSFGQARLADQGPSWASGDSSNNHKPREMLEIDLGEKKKITGIRTT 330

```

Db 304 -----WTPOOSRLHGDNGTTPNLDN-----KEYLOVDLRLFLMLTAIATQ 345
Oy 331 G--STOSNENFYKSFVNMFKNNNSKMTYKGIYVNEEKVFGNSFRDPVONNFIPIV 388
Db 346 GAISETOKGYIVKSKYLEVSTNGEDMAYYRHKNH--KIFQANDATEVVLNKLHMPFL 403
Oy 389 ARYVAVPQTMHQRALVELICQITQGNDSLVRKTS--QSTSVSTKKEDETTIRPI 445
Db 404 TRFIRPQTMHGLALRLLEFGCHVTDAPCSNMLGMLSGLIADQISASTREYLWSPS 463
Oy 446 -----PSEETSGINITVAIPVLVLYVPAKGIFA-----APRK 482
Db 464 AARLVSSSGWFPNPOAPGEEW--LOYDLGTPKTVKGILOGANGDSITAVEARAFVR 522
Oy 483 KKKKSPYGSAAEAKTDCKWKQIKYEPARHQA--EFTISYD-----521
Db 523 KFKVSYSLNGKD-----WEYIOPRTQOTKLEBGNMHTDFDIRFPDVPVPAQYRYVPE 576
Oy 522 --NEKEMTOKDLITSDMADYQOPL--MIGTGTVTRKSGTFRPMOTDAEAG 569
Db 577 RWSFAGIRLEVLGCDWTDSPKPYETELGPTVKSEETTPPYMDADAECEG 627

```

RESULT 5

```

US-08-936-135-10
: Sequence 10, Application US/08936135
: Patent No. 6054293
: GENERAL INFORMATION:
: APPLICANT: Tessier-Lavigne, Marc
: APPLICANT: He, Zhigang
: APPLICANT: Chen, Hang
: TITLE OF INVENTION: Semaphorin Receptors
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 75 DENISE DRIVE
: CITY: HILLSBOROUGH
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94010
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/936,135
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UC97-288-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 343-4341
: TELEFAX: (650) 343-4342
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 909 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-936-135-10

```

Query Match

```

9.4%: Score 354.5; DB: 3; Length 909;
Best Local Similarity 21.5%; Pred. No. 2.4e-22;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

```

```

Oy 41 CGHETTYDSTGMTSKNPGTYPNHTVEKITYVK--GKRLLRLG--DLDISQTCADY 98
Db 28 CGGRNSMDAGYITSPGIPDYPSNQCENWYVAPENQKIVLNNPFIETLEKHKDKIDF 87

```

```

Oy 99 ILF-----TSSDQYPGYCSMTVPKELLNTSEVTVRESGSHISGRGLITY-----AS 149
Db 88 IETRRGDSNADLKKHCNT--APPTIISGSGVLIKTSDARAGAGFSLRYELFKGS 146
Oy 150 SD-----HPDLITC---LERASHYKTEYSKP-----173
Db 147 EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPMHEILLQFLFDLEHDPLOVEG 206
Oy 174 -CPACGPAVDIS--GNWVQYRDTSLCKRAHAGIADLGQISVLRKGSIRY--228
Db 207 DCKYMLDIWDGIPVAVGPIGKCYCTKTPSKLRSTGILSLTFHTDMAVADGESARYL 266
Oy 229 -----EGLIANGVLSRDGSLSDKRLFTSNGCSRSLSPEPDQIRAS 270
Db 267 IHQEPENPQCVPLGMSGRIANEQISASTFSDGR-----303
Oy 271 SSMQSVNMSGQVHNSPQARLQDQGPMSASDSSNNHKKPRMWEIDYGEKKKITGITTT 330
Db 304 -----WTPOOSRLHGDNGTTPNLDN-----KEYLOVDLRLFLMLTAIATQ 345
Oy 331 G--STOSNENFYKSFVNMFKNNNSKMTYKGIYVNEEKVFGNSFRDPVONNFIPIV 388
Db 346 GAISETOKGYIVKSKYLEVSTNGEDMAYYRHKNH--KIFQANDATEVVLNKLHMPFL 403
Oy 389 ARYVAVPQTMHQRALVELICQITQGNDSLVRKTS--QSTSVSTKKEDETTIRPI 445
Db 404 TRFIRPQTMHGLALRLLEFGCHVTDAPCSNMLGMLSGLIADQISASTREYLWSPS 463
Oy 446 -----PSEETSGINITVAIPVLVLYVPAKGIFA-----APRK 482
Db 464 AARLVSSSGWFPNPOAPGEEW--LOYDLGTPKTVKGILOGANGDSITAVEARAFVR 522
Oy 483 KKKKSPYGSAAEAKTDCKWKQIKYEPARHQA--EFTISYD-----521
Db 523 KFKVSYSLNGKD-----WEYIOPRTQOTKLEBGNMHTDFDIRFPDVPVPAQYRYVPE 576
Oy 522 --NEKEMTOKDLITSDMADYQOPL--MIGTGTVTRKSGTFRPMOTDAEAG 569
Db 577 RWSFAGIRLEVLGCDWTDSPKPYETELGPTVKSEETTPPYMDADAECEG 627

```

RESULT 6

```

US-08-936-135-12
: Sequence 12, Application US/08936135
: Patent No. 6054293
: GENERAL INFORMATION:
: APPLICANT: Tessier-Lavigne, Marc
: APPLICANT: He, Zhigang
: APPLICANT: Chen, Hang
: TITLE OF INVENTION: Semaphorin Receptors
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 75 DENISE DRIVE
: CITY: HILLSBOROUGH
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94010
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/936,135
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UC97-288-2
: TELECOMMUNICATION INFORMATION:

```

```

Oy 41 CGHETTYDSTGMTSKNPGTYPNHTVEKITYVK--GKRLLRLG--DLDISQTCADY 98
Db 28 CGGRNSMDAGYITSPGIPDYPSNQCENWYVAPENQKIVLNNPFIETLEKHKDKIDF 87

```


QY 522 --NEKEMTKIDLITSDMADYQOPL-MIGTGTVRKSGTFRPMOTDAEAG 569
Db 577 RWSBAGIGMRLEVLGCDWTDKSPVETLGPVKSSETTTPYPMDEDATECG 627

RESULT 8
US-08-936-135-16
Sequence 16, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 931 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-16

Query Match
Best Local Similarity 21.5%; Score 354.5; DB 3; Length 931;
Matches 140; Conservative 96; Mismatches 242; Indels 173; Gaps 22;

QY 41 CGHLVYQDSGTMTSKNPGYTPHNTVEKTIIVPK-GKRLILRLG-DLDESQTCASDY 98
Db 28 CGGRNSKADAGYITSPGYPODYPHONCEWIVVAPEPQKIVLNFPELEKDKCKYDF 87

QY 99 LLF-----TSSDQGYPCGSMVPRKELLNTSEVTVRFESSHISGRFLTY-----AS 149
Db 88 IEIRDGSESADLLGKHGNT-APPTIISGSVYIKFTSDYARQAGAFSLRYEIKFTGS 146

QY 150 SD-----HFDLITC-----LEASHVLTKEYSK----- 173
Db 147 EDCSKNFTSPMGITSEPGFPEKRYPHNLDCTTILAKPMETIILQFLITDLEHDLQVGE 206

QY 174 -CPAGCRDVADIS--GNMVDGYRDTSLCKRAIHAGIADLGQQLSLQKGISRY-- 228
Db 207 DCKYDMDIMGIPHYGLKCYGCTKPSKLSRSTGILSLPHFDMAVANDGSARYLL 266

QY 229 -----EGLANGVLSRQGSLSDKRFLFTSNGCSLSLSTPEPQIRAS 270
Db 267 IHQEPPEFQCNVPLGMSGRIANEOQISASSTFSDGR----- 303

QY 271 SSMQSVNESGQVHMSPGQARLDQGPWASGSDSNHKKPREWLEIDLGEKKKITGRTT 330

Db 304 -----WTPQOSRLHGDDNGWTPNLDNSN---KEYIQVDRLRFLTMLTAITQ 345
QY 331 G-STQSNFNFYVASFVAFNFRNNNSKMTYKGIYVNEKEVFOGNSNFRDPYONNFIPIV 388
Db 346 GALSRETQKGYVASKYLEVSTNGEDMMVYRHGNH--KIQANNDATVEVLLNLHMLPL 403

QY 389 ARYRVVPTWTHORIALKVELIGCOITGQNDLSVWRKTS---OSTSVSTKKEDETITRPI 445
Db 404 TRFIRIRPQVTHLGIALLLEFGCRVTDAPCSNMIGMLSGLIADTQISASSTREYLSPS 463

QY 446 -----PEESTGINITVAIPVLVLVAFAGGIFA-----AFRK 482
Db 464 AARLVSSRSGWFPNPNPOAQOPEEW-LQYDLGTPKTVKCVITQAGARGDSTIAVARAVR 522

QY 483 KKKGSPYSAEAKQTDCKQIKYPFARHQA-EFTISYD----- 521
Db 523 KFKVSYSLNGMD-----WEYIQDPRTQGTLEFGNNHIDYDPDIRPDPVAPQVRYRPE 576

QY 522 --NEKEMTKIDLITSDMADYQOPL-MIGTGTVRKSGTFRPMOTDAEAG 569
Db 577 RWSBAGIGMRLEVLGCDWTDKSPVETLGPVKSSETTTPYPMDEDATECG 627

RESULT 9
US-08-936-135-18
Sequence 18, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-18

Query Match
Best Local Similarity 22.2%; Score 351.5; DB 3; Length 909;
Matches 142; Conservative 99; Mismatches 253; Indels 143; Gaps 25;

QY 41 CGHLVYQDSGTMTSKNPGYTPHNTVEKTIIVPK-GKRLILRLG-DLDESQTCASDY 98
Db 28 CGGRNSKADAGYITSPGYPODYPHONCEWIVVAPEPQKIVLNFPELEKDKCKYDF 87


```

STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 923 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-6

```

Query Match 9.0%; Score 338.5; DB 3; Length 923;

Best Local Similarity 20.8%; Pred. No. 6, Se-21;

Matches 158; Conservative 106; Mismatches 267; Indels 227; Gaps 30;

```

QY 17 RGL-----LALLIIVAPRLQAEELGDCGHLVYODSGTMSKPYGTYPHNVCCK 70
DB 3 RGLPLCATLALALALAGAF-----SDKCGGTIKIENPGYLTSPGPHSEKCEW 56
QY 71 TITVPKG-KRLILRLG-DLDESOTCASDYLFTSSSDQ-----YGPYGSMTVPKELLN 124
DB 57 LIOAPEYQRIIMFNPHFDLEDKCKYDVEYIDGENEGRLMGKFCGKI-APSPVSS 115
QY 125 TSEYTVFESGSHISGRGFLTYA-----SSDHPDLITC-- 158
DB 116 GPFLEIKFVSDYETHGAGFSIRYEIFKRGPCSONTAPTCVIKSPGPEKYPRLSCTY 175
QY 159 -----LEASHYLTKEYSKFCPAG--CRDVAAGDISGNMVDGYRDTSLCKAIIH 205
DB 176 IIFAPKXSEILIEFESFDLEDQSNP--PGGMFCRYDLEI-----WDGPEP-----YGP 223
QY 206 AG-IIADELGSOI-----SVLQRKGISREGLANGVLSRGSLSDKRFLF 250
DB 224 IGRYCGOKTPERIKSSGVLSMVYTDALAKESFSANYSVLQSSI--SEDEK----- 274
QY 251 TSNCCSRSLSEEP-----DGOIRASSSMOSVNESGDQVHWSFGQARLADOGSPMASGDSN 306
DB 275 -----CMALGMESEIHSDDOTASSOYGT-----NMYSVERSLNPEMGWTPGENDSY 322
QY 307 NHRERLELDGKKKITTGIRTTG--STOSNENFYVKSFSYFANFNKNNKSKRTYKGIYNN 364
DB 323 -----KEMIOVDLGLRVTAVGTGAIKSKETKRYKYKTYKVDISSNGEDWISLK--EGN 376
QY 365 EEKVFQNSNRPDYPVONNFIPIYARVYVPOIWHORIALKVELLICOTOGNDSLWVR 424
DB 377 KALIFQONTNTDVAVLGVESKPLITRFRVRIKPVSMETGISMRFEVYCGCKITD-----YP 430
QY 425 KTSOSTSVSTRKEDEITRPIPESETSTGINITV-----AIPVLVLLVAVFAGMGIFA 478
DB 431 CSGLGNVSGILSDQITASQADRNMMPEMIRILVTSRTGALP----- 474
QY 479 AFRKKKKSGPYGSAEAKTDCKROI-----KTPPARHOSAEFTISYDN 522
DB 475 -----PSPHY-----INEMLOVDLGDKEKIVRGVILQGGKHKREKVFARKEKFIAYSN 521
QY 523 EKEWTQALDILITSDMAYVQOPLMIGTGVTRKSGSTFRPMQIDAEAGVSTDAGHYDPOQ 582

```

```

DB 522 -----NGSDWKTIMDSRKRAKSPFGNNNYDTPPE 550
QY 583 RAGRHEVALPAPPEPEYATPI-----VERHYLRATFTFSAGSYRVPQPGKHSLSSGGF 639
DB 551 LRTSPISLRIRIYPERATHSGILRRELJGCEVEAPTAPGTPNCPNVE----- 602
QY 640 SPVAVGAGDGDYORPHSAQPADRGYDRPKAVSALATE 677
DB 603 -----CDDQANCMHSGTGDD--FOLTGTGYLATE 630

```

RESULT 12

US-09-116-473-4

Sequence 4, Application US/09116473

Patent No. 6428965

GENERAL INFORMATION:

APPLICANT: Kolodkin, Alex

APPLICANT: Ginty, David

TITLE OF INVENTION: SEMAPHORIN RECEPTOR

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff

STREET: 1001 G Street, NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/116,473

FILING DATE: 17-JUL-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/052,762

FILING DATE: 17-JUL-1998

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A

REGISTRATION NUMBER: 32141

REFERENCE/DOCKET NUMBER: 01107,74973

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 922 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-116-473-4

Query Match 8.8%; Score 331; DB 4; Length 922;

Best Local Similarity 24.1%; Pred. No. 3e-20;

Matches 112; Conservative 81; Mismatches 166; Indels 106; Gaps 19;

```

QY 17 RGL-----LALLIIVAPRLQAEELGDCGHLVYODSGTMSKPYGTYPHNVCCK 70
DB 3 RGLPLCATLALALALAGAF-----SDKCGGTIKIENPGYLTSPGPHSEKCEW 56
QY 71 TITVPKG-KRLILRLG-DLDESOTCASDYLFTSSSDQ-----YGPYGSMTVPKELLN 124
DB 57 LIOAPEYQRIIMFNPHFDLEDKCKYDVEYIDGENEGRLMGKFCGKI-APSPVSS 115
QY 125 TSEYTVFESGSHISGRGFLTYA-----SSDHPDLITC-- 158
DB 116 GPFLEIKFVSDYETHGAGFSIRYEIFKRGPCSONTAPTCVIKSPGPEKYPRLSCTY 175
QY 159 -----LEASHYLTKEYSKFCPAG--CRDVAAGDISGNMVDGYRDTSLCKAIIH 205

```

Db 176 IIFAPKSEIILEFESEDLQDSDNP--PGGVFCRYDRLEI-----WQGFPE----- 219
QY 206 AGIADBLGQISVLQKKGISREYGLA-----NGVLSRGSLSDKRFLFTSNG-----CS 256
Db 220 ---VGPPIGRCQCKTPGRIRSSGILISMVFTYDSAIKAGFSANYSVLQSSISEDKCM 276
QY 257 RSLSEFEP---DQIRASSWSQVNSGDOVHNSPGQARLQDGPSPWASGSDSSNNHRPRE 312
Db 277 EALMGESGEIHSDQITVASSQYGT-----NMSVERSLNYPENGWTPGEDSY-----RE 324
QY 313 WLEIDGEEKKKTIGRTG--STQSNFNFYKSFVNMFKNNNSKWKYKGIYVNEEVQ 370
Db 325 WIGVDGLAFYAVGTQGAISKETKKKYVKKYKRYVDISSNGEDWTLK--EGNKATIFQ 382
QY 371 GNSNFRDPVONNFIPIVARYVVPQTHQRIAKVELIGCOT 415
Db 383 GNTNPIDVYGVFPKPLITFRVRIKPSAWMETGISMREYVGCKIT 427

RESULT 13

US-07-607-538C-4

Sequence 4, Application US/07607538C

Patent No. 5455031

GENERAL INFORMATION:

APPLICANT: Ceriani Dr., Roberto L.

APPLICANT: Peterson Dr., Jerry A.

APPLICANT: Larocca, David J.

TITLE OF INVENTION: POLYPEPTIDE WITH 46

TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING

TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGUES,

TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-

TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-

TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF

TITLE OF INVENTION: USE THEREOF

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: V. Amzel & Assoc.

STREET: 2055 No. 545031th Broadway

CITY: Walnut Creek

STATE: California

COUNTRY: USA

ZIP: 94596

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/607,538C

FILING DATE: 01-NOV-1990

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Viviana Amzel

REGISTRATION NUMBER: 30,930

REFERENCE/DOCKET NUMBER: CRECC-004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 943-1931

TELEFAX: (510) 943-1189

TELEX: N.A.

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 218 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE:

US-07-607-538C-4

Query Match 8.5%; Score 321; DB 1; Length 218;
Best Local Similarity 42.6%; Pred. No. 2,1e-20;
Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;

QY 253 NCGSRSLSEF-----PDQIRASS--SWQVNSGDOVHNSPGQARLQDGPSPWASGSDSS 305
Db 58 NCGSTPLGMENKRIENKQITVASSFKSMW-----GD--YMEPRARLNAQGRVNAQAKA 110
QY 306 NNHRPREWLEIDGEEKKKTIGRTGSTQSNFNFYKSFVNMFKNNNSKWKYKGIYVNE 365
Db 111 NNHR--QWLEIDLKIKKRIITAITQCGCKSLSEMYKSYTIHYSEQGVEMKPYRLKSSMV 168
QY 366 EKVQSGNSNFRDPVONNFIPIVARYVVPQTHQRIAKVELIGQI 414
Db 169 DKIFEGNTKRGVKNFNPITISRTVLPKWNOSIALRLLEFGDI 217

RESULT 14

US-08-162-402B-4

Sequence 4, Application US/08162402B

Patent No. 5972337

GENERAL INFORMATION:

APPLICANT: CERIANI, ROBERTO L.

APPLICANT: PETERSON, JERRY A.

APPLICANT: LARocca, DAVID J.

TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT

TITLE OF INVENTION: GLOBULE (HMFg) ANTIGEN, FRAGMENTS & FUSION PROTEIN

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder & Poplawski

STREET: 444 South Flower St., 19th Floor

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/162,402B

FILING DATE: 03-DEC-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Amzel, Viviana

REGISTRATION NUMBER: 30,930

REFERENCE/DOCKET NUMBER: P66 38215

TELECOMMUNICATION INFORMATION:

TELEPHONE: 213-622-7700

TELEFAX: 213-489-4210

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 218 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-162-402B-4

Query Match 8.5%; Score 321; DB 2; Length 218;
Best Local Similarity 42.6%; Pred. No. 2,1e-20;
Matches 72; Conservative 30; Mismatches 51; Indels 16; Gaps 5;

QY 253 NCGSRSLSEF-----PDQIRASS--SWQVNSGDOVHNSPGQARLQDGPSPWASGSDSS 305
Db 58 NCGSTPLGMENKRIENKQITVASSFKSMW-----GD--YMEPRARLNAQGRVNAQAKA 110
QY 306 NNHRPREWLEIDGEEKKKTIGRTGSTQSNFNFYKSFVNMFKNNNSKWKYKGIYVNE 365
Db 111 NNHR--QWLEIDLKIKKRIITAITQCGCKSLSEMYKSYTIHYSEQGVEMKPYRLKSSMV 168

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:18:56 ; Search time 36.0911 Seconds

(without alignments)
2386.400 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDGCGHLTYQDSGTWT.....STKEDETTRPPEESTST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 504360

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	6.7	432	11	Q8R327 mus musculu
2	28	6.7	460	11	Q9D696 mus musculu
3	28	6.7	503	11	Q9D4J3 mus musculu
4	9	2.2	440	16	Q8X543 escherichia
5	8	1.9	216	4	Q96H40 homo sapien
6	8	1.9	255	16	Q9ZG35 rickettsia
7	8	1.9	264	16	Q8YPN2 Oryz235 ricke
8	8	1.9	339	10	Q9XEX1 Oryxex1 gracila
9	8	1.9	364	4	Q9H8G1 Oryh8g1 homo
10	8	1.9	470	4	Q9B8E6 Oryb8e6 homo
11	8	1.9	519	4	Q96K04 Ory6k04 homo
12	8	1.9	524	4	Q96U04 Ory6u04 homo
13	8	1.9	536	4	Q96B85 Ory6b85 homo
14	8	1.9	612	10	Q9AUC3 Oryauc3 lycoper
15	8	1.9	657	4	Q96DM2 Ory6dm2 homo
16	7	1.7	100	16	Q9PCB6 Orypcb6 xylella

17	7	1.7	111	10	Q8VX51 O8vx51 cicer arlet
18	7	1.7	159	13	Q9DEQ9 Q9deg9 gallus gall
19	7	1.7	159	17	Q8TTR5 Q8ttr5 methanosarc
20	7	1.7	162	9	Q94M38 Q94m38 streptococc
21	7	1.7	182	5	Q9UAX2 Q9uax2 caenorhabd
22	7	1.7	183	17	Q9HSH4 Q9hsh4 halobacteri
23	7	1.7	213	16	Q97M44 Q97m44 clostridium
24	7	1.7	230	10	Q947Y3 Q947y3 oryza sativ
25	7	1.7	231	5	Q9TZ96 Q9tze96 caenorhabd
26	7	1.7	233	5	Q9VLE1 Q9vle1 drosophila
27	7	1.7	237	11	Q9CUV2 Q9cuv2 mus musculu
28	7	1.7	242	2	Q9KRB9 Q9krb9 ervania chr
29	7	1.7	254	5	Q21223 Q21223 caenorhabd
30	7	1.7	256	5	Q8SQ06 Q8sq06 encaphalito
31	7	1.7	260	16	Q8X4D9 Q8x4d9 escherichia
32	7	1.7	261	16	Q9KEX1 Q9kex1 bacillus ha
33	7	1.7	265	10	Q41069 Q41069 oryza sativ
34	7	1.7	270	7	Q98018 Q98018 hylobates l
35	7	1.7	270	7	Q98026 Q98026 cercopithec
36	7	1.7	272	7	Q98024 Q98024 cercopithec
37	7	1.7	272	7	Q98025 Q98025 papio sp. (
38	7	1.7	272	10	Q80422 Q80422 oryza sativ
39	7	1.7	272	16	Q98OV3 Q98ov3 mycoplasma
40	7	1.7	273	10	Q40628 Q40628 oryza sativ
41	7	1.7	275	7	Q98020 Q98020 pan paniscu
42	7	1.7	275	7	Q98023 Q98023 cercopithec
43	7	1.7	285	16	Q9KAK8 Q9kak8 bacillus ha
44	7	1.7	286	16	Q8ZOH9 Q8zoh9 salmonella
45	7	1.7	293	16	Q9CKB0 Q9ckb0 pasteurella

ALIGNMENTS

RESULT 1
ID Q8R327 PRELIMINARY; PRT; 432 AA.
AC Q8R327;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to RIKEN CDNA 4631413k11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026771; AAH26771.1; -
SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 6.7%; Score 28; DB 11; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.4e-21;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 GSGHISGRGFLTYASSDHPDLITCLER 127
DB 54 GSGHISGRGFLTYASSDHPDLITCLER 81
|||||
RESULT 2
ID Q9D696 PRELIMINARY; PRT; 460 AA.
AC Q9D696;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 4631413K11Rik protein.
GN 4631413K11Rik
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AK014521; BAB29409.1; -
DR MGD: MGI:1913936; 463143K11R1K.
DR InterPro: IPR000859; CUB_domain.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR PROSITE: PS01180; CUB; 1.
SQ SEQUENCE 460 AA; 50334 MW; 3CF3556F70A88B93 CRC64;

Query Match 6.7%; Score 28; DB 11; Length 460;
Best Local Similarity 100.0%; Pred. No. 4.7e-21;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 SCSHISGRFLTYASSDHPDLITCLER 127
DB 82 SCSHISGRFLTYASSDHPDLITCLER 109
|||||
ID 09D4J3 PRELIMINARY; PRT; 503 AA.
AC 09D4J3;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE 463143K11R1K protein.
GN 463143K11R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AK014521; BAB29409.1; -
DR MGD: MGI:1913936; 463143K11R1K.
DR InterPro: IPR000859; CUB_domain.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR PROSITE: PS01180; CUB; 1.
SQ SEQUENCE 503 AA; 54547 MW; FDB121E845CA068 CRC64;
```

```
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AK014521; BAB29409.1; -
DR MGD: MGI:1913936; 463143K11R1K.
DR InterPro: IPR000859; CUB_domain.
DR Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR PROSITE: PS01180; CUB; 1.
SQ SEQUENCE 503 AA; 54547 MW; FDB121E845CA068 CRC64;

Query Match 6.7%; Score 28; DB 11; Length 503;
Best Local Similarity 100.0%; Pred. No. 5.1e-21;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 SCSHISGRFLTYASSDHPDLITCLER 127
DB 125 SCSHISGRFLTYASSDHPDLITCLER 152
|||||
ID 08X543 PRELIMINARY; PRT; 440 AA.
AC 08X543;
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Regulator of uhpT.
GN UHPC OR 25157 OR EGS4604.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;
RL Nature 409:529-533(2001).
RA [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11256796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT *Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.*;
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005598; AAG58866.1; -
DR MGD: MGI:1913936; 463143K11R1K.
DR InterPro: IPR000849; G1PT_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRFAMs: TIGR00881; ZAO104; 1.
DR PROSITE: PS00942; G1PT; 1.
KW Complete proteome.
SQ SEQUENCE 440 AA; 48345 MW; C3442BAC1132448 CRC64;
```

Query Match 2.2%; Score 9; DB 16; Length 440;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
 |||||
 DB 57 ILANGVLSR 65

RESULT 5

ID 096H40 PRELIMINARY; PRT; 216 AA.
 AC 096H40;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Similar to DNA-binding protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRISUB-LONG;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC008936; AAH08936.1; -
 DR InterPro: IPR001909; KRA8.
 DR InterPro: IPR000822; znf_C2H2.
 DR Pfam: PF01352; KRA8; 1.
 DR PROSITE: PS50805; KRA8; 1.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 2.
 DR DNA-binding; Zinc-finger.
 KW SEQUENCE 216 AA; 25227 MW; B5C6CF182AFEC2 CRC64;
 SQ

Query Match 1.9%; Score 8; DB 4; Length 216;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 50 PDLITCLE 57

RESULT 6

ID 092G35 PRELIMINARY; PRT; 255 AA.
 AC 092G35;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical protein R01290.
 GN R01290.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MALISH 7;
 RX MDLME-21442074; PubMed-11557893;
 RA Ogata H., Audic S., Remesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Reaout D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098(2001).
 DR EMBL: AE008676; AAU03828.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 255 AA; 29018 MW; 4E4A5867E03C9B3 CRC64;
 Query Match 1.9%; Score 8; DB 16; Length 255;
 Best Local Similarity 100.0%; Pred. No. 9.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 370 ALKVELIG 377
 |||||
 DB 242 ALKVELIG 249

RESULT 7

ID 08YPN2 PRELIMINARY; PRT; 264 AA.
 AC 08YPN2;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein Alr4161.
 GN ALR4161.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-21595285; PubMed-11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL: AP003595; BAB75860.1; -
 DR InterPro: IPR001989; Radical_activat.
 DR Pfam: PF02143; Radical_activat; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 264 AA; 29421 MW; BDF27F3428A007D0 CRC64;

Query Match 1.9%; Score 8; DB 16; Length 264;
 Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 357 YRVVPQT 364
 |||||
 DB 250 YRVVPQT 257

RESULT 8

ID 09XEX1 PRELIMINARY; PRT; 339 AA.
 AC 09XEX1;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Trypophan synthase (EC 4.2.1.20) (Fragment).
 OS Gracilariaria veruicosa.
 OC Eukaryota; Rhodophyta; Florideophyceae; Gracilariiales; Gracilariaceae;
 OC Gracilariia.
 NCBI_TaxID=2777;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lulisma A.O., Ragan M.A.;
 RT "Occurrence of closely spaced genes in the nuclear genome of the
 agarophyte Gracilariaria gracilis.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L-SERINE + 1-(1NDOL-3-YL)GLYCEROL 3-PHOSPHATE
 CC = L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 CC -1- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
 CC -1- SIMILARITY: TO THE TRPB FAMILY.
 DR EMBL: AF121272; AAD17312.1; -
 DR HSSP: P00933; 2MSY
 DR InterPro: IPR001926; B6_enzyme_beta.
 DR InterPro: IPR000993; Trp_synth_beta.
 DR Pfam: PF00291; PALP; 1.

DR PROSITE: PS00168; TRP-SYNTASE_BETA; 1.
 KM Lyase; Pyridoxal phosphate; tryptophan biosynthesis.
 FT NON_TER 339
 SQ SEQUENCE 339 AA; 37422 MW; 89ADAF872162349 CRC64;

Query Match 1.9%; Score 8; DB 10; Length 339;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 394 OSTSVSTR 401
 |||||
 DB 2 OSTSVSTR 9

RESULT 9

O9H8G1 PRELIMINARY: PRT: 364 AA.

AC O9H8G1:
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 21, Last annotation update)
 DE CDNA FLJ13659 f18, clone PLACE1011576, moderately similar to human
 DE kruppel related zinc finger protein (Htf10) mRNA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACEMENT:
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
 RA Watanabe S., Hosokawa T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.,
 RT "NEBO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK033721; BAB14656.1; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 2.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 KM DNA-binding; Metal-binding; Zinc-finger.
 RN SEQUENCE 364 AA; 42472 MW; 5043DA4C575842B0 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 364;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 82 PDLITCLE 89

RESULT 10

O9B2E6 PRELIMINARY: PRT: 470 AA.

AC O9B2E6:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE KRAB zinc finger protein HZF26 (Fragment).
 GN HZF26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-95169271; PubMed-7865130;
 RA Ahrink M., Aveskog M., Hellman L.;
 RT "Isolation of cDNA clones for 42 different Kruppel-related zinc finger
 RT proteins expressed in the human monoblast cell line U-937."
 RL DNA Cell Biol. 14:125-136(1995).
 RN [2]

RP SEQUENCE FROM N.A.

RA Ahrink M., Aveskog M., Hellman L.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL; AF325191; AA01422.1; -
 DR HSSP; P08046; 1AIG.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf-C2H2; 10.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 9.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 11.
 KM DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
 RN NON_TER 470 470
 SQ SEQUENCE 470 AA; 54961 MW; 997CB1C95D37E983 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 470;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||
 DB 50 PDLITCLE 57

RESULT 11

O96M04 PRELIMINARY: PRT: 519 AA.

AC O96M04:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CDNA FLJ32933 f18, clone TEST12007466, moderately similar to zinc
 DE finger protein 91.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE-TESTIS:

RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Hotuta T., Hiraoaka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Watanabe M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
 RT "NEBO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK057495; BAB1510.1; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRAB; 1.
 DR Pfam; PF00096; Zf-C2H2; 12.
 DR PROSITE; PS00805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN; 8.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 12.

KW DNA-binding; Zinc-finger.
SQ SEQUENCE 519 AA; 60350 MW; 7E4D17E53EC7732C CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 519;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
|||||||

Db 59 PDLITCLE 66

RESULT 12
Q96JCA PRELIMINARY; PRT; 524 AA.

AC Q96JCA; 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE KRAB zinc finger protein.
GN KR19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE=21303228; PubMed=11410164;
RA Mark C., Looman C., Ahrink M., Hellman L.;
RT Molecular cloning and preliminary functional analysis of two novel
RT human KRAB zinc finger proteins, HKR18 and HKR19.";
RL DNA Cell Biol. 20:275-286(2001).
DR EMBL: AF277624; AAK61307.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000832; Znf_C2H2.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; zf_C2H2; 12.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_11.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 12.
DR KW DNA-binding; zinc-finger
SQ SEQUENCE 524 AA; 60598 MW; F1AD4929DC67D105 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 524;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
|||||||

Db 62 PDLITCLE 69

RESULT 13
Q96BB5 PRELIMINARY; PRT; 536 AA.

AC Q96BB5; 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical 61.9 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-B-CELL;
RA Strausberg R.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC015765; AAH15765.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.

DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; zf_C2H2; 13.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 13
KW DNA-binding; Hypothetical protein; Zinc-finger.
SQ SEQUENCE 536 AA; 61932 MW; 3DC80FAE53F8CC00 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 536;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
|||||||

Db 50 PDLITCLE 57

RESULT 14
Q9AUC3 PRELIMINARY; PRT; 612 AA.

AC Q9AUC3; 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Receptor-like protein kinase 3.
GN PRK3.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanales; Solanaceae; Solanum.
OX NCBI_Taxid=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. VF36;
RA Kim H.-U., Cotter R., McCormick S.;
RT Arabidopsis: the tomato kinases identified in tomato, maize and
RT expression patterns during pollen tube growth.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF243040; AAK28345.1; -
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00560; LRR; 5.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00370; LRR; 5.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; TYKc; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 612 AA; 68096 MW; A39B3C3751C3F27 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 612;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 KELLANTS 92
|||||||

Db 31 KELLANTS 38

RESULT 15
Q96DW2 PRELIMINARY; PRT; 657 AA.

AC Q96DW2; 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Similar to RecQ protein-like 4 (Fragment).
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LYMPH;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013277; AAH13277.1; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
KW ATP-binding; Helicase.
FT
FT NON_TER 1 1
SQ SEQUENCE 657 AA; 73415 MW; CF51B910F560CA18 CRC64;

Query Match 1.9%; Score 8; DB 4; Length 657;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 PGQARLQD 260
Db 576 PGQARLQD 583

Search completed: May 15, 2003, 13:24:08
Job time : 40.0911 secs

PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -

PS Claim 5; Page 29-30; 33pp; English.

CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.

XX Sequence 539 AA;

Query Match 100.0%; Score 418; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 60
DB 35 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 94
QY 61 ASDYLLFTSSSDQYGYCGSMTPVPEKLLNTSEVTVRESGSHISGRGFLITYASSDHPD 120
DB 95 ASDYLLFTSSSDQYGYCGSMTPVPEKLLNTSEVTVRESGSHISGRGFLITYASSDHPD 154
QY 121 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDGYRDTSLCKRAIHAGITADBLG 180
DB 155 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDGYRDTSLCKRAIHAGITADBLG 214
QY 181 GOISVLOQRKGISREGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
DB 215 GOISVLOQRKGISREGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 274
QY 241 SVNBSGDVHMSPGQARLQDGPMSWASGSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 300
DB 275 SVNBSGDVHMSPGQARLQDGPMSWASGSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 334
QY 301 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVQNNFIPPIVARYRV 360
DB 335 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVQNNFIPPIVARYRV 394
QY 361 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 418
DB 395 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 452

RESULT 2
AAU00629
ID AAU00629 standard; Protein: 586 AA.

XX AAU00629;

XX 29-AUG-2001 (first entry)

DE Novel human protein (NHP) sequence #2.

KM Novel human protein: NHP; CUB domain; extracellular domain; gene therapy;

KM obesity; high blood pressure; connective tissue disorder; infertility;

XX NHP-mediated pathway.

XX Homo sapiens.

PN WO200129219-A1.

XX 26-APR-2001.

XX 08-OCT-2000; 2000WO-US28798.

XX 19-OCT-1999; 99US-0160285.

PR 18-FEB-2000; 2000US-0183583.

XX (LEXI-) LEXICON GENETICS INC.

XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX WPI; 2001-290917/30.

DR N-PSDB; AAS00614.

XX Claim 2; Page 27-28; 33pp; English.

CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.

XX Sequence 586 AA;

Query Match 100.0%; Score 418; DB 22; Length 586;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 60
DB 82 EELGDCGHLVLYOISGTMSTKNPGTYPNHTVCEKTTVPKGRLLIRLGLDIESQTC 141
QY 61 ASDYLLFTSSSDQYGYCGSMTPVPEKLLNTSEVTVRESGSHISGRGFLITYASSDHPD 120
DB 142 ASDYLLFTSSSDQYGYCGSMTPVPEKLLNTSEVTVRESGSHISGRGFLITYASSDHPD 201
QY 121 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDGYRDTSLCKRAIHAGITADBLG 180
DB 202 LITCERASHLKTETSEFCPCAGCDVAGDISGNMVDGYRDTSLCKRAIHAGITADBLG 261
QY 181 GOISVLOQRKGISREGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 240
DB 262 GOISVLOQRKGISREGLIANGVLSRDGSLSDKRFLETSNGCSRSLSFEPDQIRASSSMQ 321
QY 241 SVNBSGDVHMSPGQARLQDGPMSWASGSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 300
DB 322 SVNBSGDVHMSPGQARLQDGPMSWASGSSNNHPRFMLEIDLEGKKKITGIRTTGSTQ 381
QY 301 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVQNNFIPPIVARYRV 360
DB 382 SNFNFTYKSFYMNFRKNNNSKKTKYKGIYNNNEKVFQGSNFRDPVQNNFIPPIVARYRV 441
QY 361 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 418
DB 442 VPQTHQRIALVELIGQITQGNDSLVRKRTSOSTSVTKKEDETITRPIPSEETST 499


```

RESULT 3
AAU00628
ID AAU00628 standard; Protein: 487 AA.
XX
AC AAU00628;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #1.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN MO200129219-A1.
XX
PD 26-APR-2001.
XX
PE 08-OCT-2000; 2000MO-US28798.
XX
PR 19-OCT-1999; 990US-0160285.
XX
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
XX DR N-PSDB; AAS00613.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 4; Page 26; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
XX
SO Sequence 487 AA:
Query Match 95.7%; Score 400; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 19 MTSNNGPTGPNHVCCKTTPVGRKRLIRLGDLDIESOTGASDYLLFTSSSDQVGYC 78
DB 1 MTSKNFGITPNTKVTCKTITVGRKRLIRLGDLDIESOTGASDYLLFTSSSDQVGYC 60
OY 79 GSNTPVELLLNTSEVTRESGSHISGRFLTYASSDHPDLITCERASHYLKTEYSK 138
DB 61 GSNTPVELLLNTSEVTRESGSHISGRFLTYASSDHPDLITCERASHYLKTEYSK 120
OY 139 FCRPGCDVAGDISGNMVDGYRDTSLCKAAIHAGITADELGGISVLYQRKGISRYGIL 198
DB 121 FCRPGCDVAGDISGNMVDGYRDTSLCKAAIHAGITADELGGISVLYQRKGISRYGIL 180

```

```

OY 199 ANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSMOSVNESGDVHNSPGOARL 258
DB 181 ANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGOIRASSSMOSVNESGDVHNSPGOARL 240
OY 259 QDGPSPWASGDSNNHPRFWELEIDLGKKRTGIRTTGTSOSNFNFVYSPVNFKNNN 318
DB 241 QDGPSPWASGDSNNHPRFWELEIDLGKKRTGIRTTGTSOSNFNFVYSPVNFKNNN 300
OY 319 SKMRTYGIYNNKEKVFQGSNFPDPVQNNFIPPIVARYRVVQPTWHORIALKVELIGC 378
DB 301 SKMRTYGIYNNKEKVFQGSNFPDPVQNNFIPPIVARYRVVQPTWHORIALKVELIGC 360
OY 379 QITGNDSLVWRKTSOSTSVTKKEDETITRPISSEST 418
DB 361 QITGNDSLVWRKTSOSTSVTKKEDETITRPISSEST 400

RESULT 4
AAE22715
ID AAE22715 standard; Protein: 398 AA.
XX
AC AAE22715;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy1 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory disease; gene therapy;
KW neurotropic; neuroprotective; vulnerrary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; vitruicide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX
PN MO200222815-A1.
XX
PD 21-MAR-2002.
XX
PE 12-SEP-2001; 2001MO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
XX PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX WPI; 2002-393966/42.
XX DR N-PSDB; AAD35992.
XX
PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects -
XX
PS Claim 3; Page 123-125; 152pp; English.
XX
CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.

```

CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopenia, platelet anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy1
 CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

SO Sequence 398 AA;

Query Match 27.8%; Score 116; DB 23; Length 398;
 Best Local Similarity 100.0%; Pred. No. 2e-106;
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGGCGHLYTQDSGTMNTSKNYPGTPNHTVCEKTLTPKRGKLLIRLGDIDIESQTCASD 63
 DB 4 GGGCGHLYTQDSGTMNTSKNYPGTPNHTVCEKTLTPKRGKLLIRLGDIDIESQTCASD 63

OY 64 YLLFTSSDQYPCYCGSMTPPEKLLNLTSEVTRESGSHISGRGFLTYASSDHP 119
 DB 64 YLLFTSSDQYPCYCGSMTPPEKLLNLTSEVTRESGSHISGRGFLTYASSDHP 119

RESULT 5
 AAU79459
 ID AAU79459 standard; Protein: 398 AA.

AC AAU79459;

DT 15-JUL-2002 (first entry)

DE Human Neuropilin-Hy1.

XX Human; neuropilin-Hy1; chromosome 6q21; neuronal growth;

KW nerve regeneration; neurodegenerative disease; learning disorder;

KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;

KW organ growth; nervous system lesion; cancer; cell proliferation;

KW cell differentiation; stem cell growth factor activity;

KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;

KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;

KW platelet disorder; thrombocytopenia; liver fibrosis; lung fibrosis;

KW reperfusion; food supplement; DNA microarray.

XX Homo sapiens.

OS WO200222780-A2.

PN 21-MAR-2002.

PD 11-SEP-2001; 2001WO-US28590.

PF 11-SEP-2000; 2000US-0659671.

PR 06-SEP-2001; 2001US-0659671.

XX (TANG/) TANG T Y.

XX Tang TX;

XX WPI, 2002-351881/38.

DR N-PSDB; ABR49565.
 XX New neuropilin-like polypeptides for diagnosing, preventing and
 PT treating neurological conditions and disorders, cancers, and for
 PT inducing angiogenesis and neovascularisation
 XX Claim 3; Page 118-120; 144pp; English.

XX The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC mapping genetic neuronal defects and degenerative diseases like
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischemic lesions, infectious lesions,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for
 CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hy1.

SO Sequence 398 AA;

Query Match 27.8%; Score 116; DB 23; Length 398;
 Best Local Similarity 100.0%; Pred. No. 2e-106;
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GGGCGHLYTQDSGTMNTSKNYPGTPNHTVCEKTLTPKRGKLLIRLGDIDIESQTCASD 63
 DB 4 GGGCGHLYTQDSGTMNTSKNYPGTPNHTVCEKTLTPKRGKLLIRLGDIDIESQTCASD 63

OY 64 YLLFTSSDQYPCYCGSMTPPEKLLNLTSEVTRESGSHISGRGFLTYASSDHP 119
 DB 64 YLLFTSSDQYPCYCGSMTPPEKLLNLTSEVTRESGSHISGRGFLTYASSDHP 119

RESULT 6
 AAB19126
 ID AAB19126 standard; Protein: 503 AA.

AC AAB19126;

DT 19-FEB-2001 (first entry)

DE Polypeptide isolated from lymph node stromal cells of fsn -/- mice.

XX Lymph node stromal cell; fsn -/- mice; inflammatory disorder;

KW immune system disorder; cancer; viral disorder; HIV infection;

KW blood vessel growth; tumour necrosis factor disorder; arthritis;

KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;

KW cardiac failure.

XX Mus sp.

OS

XX WO200058463-A1.
 PN 05-OCT-2000.
 PD 18-FEB-2000; 2000MO-NZ00015.
 PF 25-MAR-1999; 9905-0276268.
 PR 26-AUG-1999; 9905-0383586.
 XX
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble RD;
 PI Murison JG;
 PI WPI; 2000-664924/64.
 DR N-PSDB; AAA96736.
 XX
 XX Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
 PT useful for modulating growth of blood cells, for treating inflammatory
 PT and tumor necrosis factor-mediated disorders, cancer and viral
 PT disorders -
 XX
 PS Claim 1; Page 68-69; 75pp; English.
 XX
 CC The present sequence represents a polypeptide sequence which is
 CC isolated from lymph node stromal cells of fsn -/- mice. The
 CC polynucleotides and their polypeptides are useful for treating an
 CC inflammatory disorder, disorder of immune system and cancer selected
 CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
 CC viral disorder, in particular HIV infection and for modulating the
 CC growth of blood vessels. The polypeptides are useful for treating a
 CC tumor necrosis factor (TNF) mediated disorder, such as those selected
 CC from arthritis, inflammatory bowel disease and cardiac failure and a
 CC fibroblast growth factor-mediated disorder. It is also useful in assays
 CC to determine biological activity, to raise antibodies, to isolate
 CC corresponding ligands or receptors, to quantify levels of protein or
 CC cognate corresponding ligand or receptors, as antiinflammatory agents,
 CC and in compositions for the treatment of skin, connective tissue and
 CC immune system diseases. The polynucleotide is useful as marker for
 CC tissue, as a chromosome marker or tags in the identification of a
 CC genetic disorder.
 CC
 CC Sequence 503 AA:
 SQ
 Query Match 6.7%; Score 28; DB 21; Length 503;
 Best local Similarity 100.0%; Pred. No. 6.8e-19;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 100 SGGSHISGRGFLTYASSDHPDLITCLER 127
 DB 125 SGGSHISGRGFLTYASSDHPDLITCLER 152
 II
 RESULT 7
 AA019467
 ID AAU19467 standard; Protein; 101 AA.
 AC AAU19467;
 XX
 XX 04-DEC-2001 (first entry)
 DT
 XX Human diagnostic and therapeutic polypeptide (DITHP) #53.
 DE
 XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder.
 XX Homo sapiens.
 OS
 XX WO200162927-A2.
 PN

PD 30-AUG-2001.
 XX
 PF 21-FEB-2001; 2001MO-US06059.
 XX
 PR 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184768.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184774.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 24-FEB-2000; 2000US-0185216.
 PR 12-MAY-2000; 2000US-0203785.
 PR 15-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'Sa SA, Amshery S, Dahl CR, Dam TC, Dattels SE;
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Dafo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
 XX
 DR WPI; 2001-502867/55.
 DR N-PSDB; AAS31038.
 XX
 PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
 PT
 XX
 PS Claim 27; Page 430; 522pp; English.
 XX
 CC The invention relates to polynucleotides (I) encoding diagnostic and
 CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
 CC and proteins involved in growth and development and receptors (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate DITHP expression. For example, (I) and
 CC (II) may be used to treat disorders associated with decreased polypeptide
 CC expression by rectifying mutations or deletions in a patient's genome,
 CC that affect the activity of the DITHPs, by expressing inactive proteins
 CC or supplementing the patient's own production of them. (I) and (II)
 CC may be used to treat diseases, for example, cell proliferative disorder,
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
 CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
 CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. (II) may also be used as antigens in the production of

CC antibodies against DITHP and in assays to identify modulators of DITHP
 CC expression and activity. The anti-DITHP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DITHP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DITHP in samples (e.g. by enzyme linked immunosorbent
 CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
 CC therapeutic (DITHP) polypeptides of the invention.

XX
 CC
 SQ Sequence 101 AA;

Query Match 1.9%; Score 8; DB 22; Length 101;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||
 Db 59 PDLITCLE 66

RESULT 8
 ABP51369
 ID ABP51369 standard; Protein: 101 AA.

XX
 AC ABP51369;

XX
 DT 03-SEP-2002 (first entry)

XX
 DE Human MDDT SEQ ID NO 391.

XX
 KW Human; MDDT: disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antinflammatory; antipsoriatic; cytosatic; anti-HIV;
 KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antiout;
 KW neuroprotective; antirheumatic; antiarthritic.

XX
 OS Homo sapiens.

XX
 PN WO200240715-A2.

XX
 PD 23-MAY-2002.

XX
 PF 06-SEP-2001; 2001WO-US27628.

XX
 PR 06-SEP-2000; 2000US-230505P.

XX
 PR 06-SEP-2000; 2000US-230514P.

XX
 PR 06-SEP-2000; 2000US-230515P.

XX
 PR 06-SEP-2000; 2000US-230517P.

XX
 PR 06-SEP-2000; 2000US-230518P.

XX
 PR 06-SEP-2000; 2000US-230519P.

XX
 PR 06-SEP-2000; 2000US-230595P.

XX
 PR 06-SEP-2000; 2000US-230597P.

XX
 PR 06-SEP-2000; 2000US-230598P.

XX
 PR 06-SEP-2000; 2000US-230599P.

XX
 PR 06-SEP-2000; 2000US-230610P.

XX
 PR 06-SEP-2000; 2000US-230653P.

XX
 PR 06-SEP-2000; 2000US-230988P.

XX
 PR 07-SEP-2000; 2000US-230989P.

XX
 PR 07-SEP-2000; 2000US-230991P.

XX
 PR 07-SEP-2000; 2000US-231163P.

XX
 PR 07-SEP-2000; 2000US-231167P.

XX
 PA (INCY-) INCYTE GENOMICS INC.

XX
 PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL,
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR,
 PI Montiyama MG, Bradley DL, Komaygi SD, Harris B, Roseberry AM,
 PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Datto A,
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR;

XX
 DR WPI: 2002-527544/56.
 DR N-PSDB: ABQ72586.

XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS

XX
 PS Claim 14; Page 541-542; 618pp; English.

XX
 CC The invention relates to an isolated human disease detection and
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (I) or a biologically active or
 CC immunogenic fragment of (I). (I) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (I) or modulates the activity of (I), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.

XX
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDDT.

XX
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDDT, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knock humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.

XX
 SQ Sequence 101 AA;

Query Match 1.9%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||
 Db 59 PDLITCLE 66

RESULT 9
 ABP51461
 ID ABP51461 standard; Protein: 101 AA.

XX
 AC ABP51461;

XX
 DT 03-SEP-2002 (first entry)

XX
 DE Human MDDT SEQ ID NO 483.

XX
 KW Human; MDDT: disease detection and treatment molecule polynucleotide;
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatotropic; antinflammatory; antipsoriatic; cytosatic; anti-HIV;
 KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antiout;
 KW neuroprotective; antirheumatic; antiarthritic.

XX
 OS Homo sapiens.

XX
 PN WO200240715-A2.

PD 23-MAY-2002.
 XX 06-SEP-2001; 2001WO-US27628.
 PF
 XX 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230585P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230865P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 XX
 PA (INCYTE GENOMICS INC.
 PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR,
 PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
 XX
 DR WPI: 2002-527544/56.
 DR N-PSDB; ABQ72678;
 XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/Inflammatory disorder
 PT e.g. AIDS
 PT
 PS
 XX Claim 14; Page 603; 618pp; English.
 XX
 CC The invention relates to an isolated human disease detection and
 CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the
 CC specification, a naturally occurring polypeptide comprising a sequence
 CC having at least 90% identity to (i) or a biologically active or
 CC immunogenic fragment of (i) (i) is useful for screening a compound for
 CC effectiveness as an agonist or antagonist, for screening a compound that
 CC specifically binds (i) or modulates the activity of (i), and for
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
 CC screening a compound for effectiveness in altering expression of a target
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
 CC detecting MDP in a sample or for assessing toxicity of a test compound,
 CC in a diagnostic test for a condition or a disease associated with the
 CC expression of MDP in a biological sample, for detecting (I) in a sample,
 CC and for purifying (I) from a sample. A composition comprising (I), an
 CC agonist or antagonist is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional MDP.
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders
 CC associated with aberrant expression of MDP, where the disorders are
 CC selected from a cell proliferative disorder such as arteriosclerosis,
 CC cirrhosis, hepatitis, psoriasis, and cancer and an
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
 CC allergy, anaemia, asthma, arterosclerosis, gout, multiple sclerosis or
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised
 CC animals or transgenic animals to model human diseases, in somatic or
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell
 CC type, for detecting differences in the chromosomal location due to
 CC translocation or inversion among normal, carrier or affected individuals
 CC and as hybridisation probes for mapping naturally occurring genomic
 CC sequences.
 CC
 SQ Sequence 101 AA:

Query Match 1.9%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 119 PDLITCLE 126
 |||||
 Db 59 PDLITCLE 66
 RESULT 10
 AAM38946
 ID AAM38946 standard; Protein; 109 AA.
 XX
 AC AAM38946;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2091.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AA158102.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 PS
 XX Example 3; SEQ ID NO 2091; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification.
 XX
 SQ Sequence 109 AA;
 Query Match 1.9%; Score 8; DB 22; Length 109;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 119 PDLITCLE 126
 |||||
 Db 82 PDLITCLE 89

RESULT 11
 AAB92967
 ID AAB92967 standard; Protein: 117 AA.
 XX
 AC AAB92967;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:11667.
 XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8: SEQ ID 11667; 2537PP + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 117 AA;
 Query Match 1.9%; Score 8; DB 22; Length 117;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 119 PDLITCLE 126
 |||||
 Db 81 PDLITCLE 88

RESULT 12
 ABP51401
 ID ABP51401 standard; Protein: 117 AA.
 XX
 AC ABP51401;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT SEQ ID NO 423.
 XX
 KM Human; MDDT; disease detection and treatment molecule polynucleotide;
 KM proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KM autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KM rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerosis;
 KM hepatotropic; antiinflammatory; antiproliferative; cytostatic; anti-HIV;
 KM antiallergic; antianemic; antiasthmatic; antiatherosclerotic; anti gout;
 KM neuroprotective; antineumatic; antiarthritic.
 XX
 OS Homo sapiens.
 XX
 PN WO200240715-A2.
 XX
 PD 23-MAY-2002.
 XX
 PF 06-SEP-2001; 2001WO-US27628.
 XX
 PR 06-SEP-2000; 2000US-230505P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230595P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230598P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230615P.
 PR 06-SEP-2000; 2000US-230865P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231167P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JT, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
 PI Momiya MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
 XX
 DR WPI: 2002-527544/56.
 DR N-PSDB: ABQ72618.
 XX
 PT Novel human disease detection and treatment polypeptide, useful in
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
 PT e.g. AIDS -

XX Claim 14; Page 562; 618pp; English.

XX The invention relates to an isolated human disease detection and

CC treatment (MDMT) polypeptide (I) selected from a polypeptide having a

CC sequence selected from 354 sequences (ABP51231-ABP51484) given in the

CC specification, a naturally occurring polypeptide comprising a sequence

CC having at least 90% identity to (I) or a biologically active or

CC immunogenic fragment of (I). (I) is useful for screening a compound for

CC effectiveness as an agonist or antagonist, for screening a compound that

CC specifically binds (I) or modulates the activity of (I), and for

CC preparing a polyclonal or monoclonal antibody by hybridoma technology.

CC Nucleic acids (II) (ABQ7449-ABQ72700) encoding (I) are useful for

CC screening a compound for effectiveness in altering expression of a target

CC polynucleotide comprising. Oligonucleotides and antibodies are useful for

CC detecting MDMT in a sample or for assessing toxicity of a test compound,

CC in a diagnostic test for a condition or a disease associated with the.

CC expression of MDMT in a biological sample, for detecting (I) in a sample,

CC and for purifying (I) from a sample. A composition comprising (I), an

CC agonist or antagonist is useful for treating a disease or condition

CC associated with decreased or increased expression of functional MDMT.

CC (I) or (II) are useful for diagnosing, treating or preventing disorders

CC associated with aberrant expression of MDMT, where the disorders are

CC selected from a cell proliferative disorder such as arteriosclerosis,

CC cirrhosis, hepatitis, psoriasis, and cancer and an

CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,

CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or

CC rheumatoid arthritis. (II) are useful for creating knockin humanised

CC animals or transgenic animals to model human diseases, in somatic or

CC germ-line gene therapy, to generate a transcript image of a tissue or cell

CC type, for detecting differences in the chromosomal location due to

CC translocation or inversion among normal, carrier or affected individuals

CC and as hybridisation probes for mapping naturally occurring genomic

CC sequences.

XX

SO Sequence 117 AA;

Query Match 1.9%; Score 8; DB 23; Length 117;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126

Db 90 PDLITCLE 97

RESULT 13

AAM90976

ID AAM90976 standard; Protein; 126 AA.

XX

AC AAM90976;

XX

DT 07-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen SEQ ID NO:18569.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytosolic; gene therapy; vaccine; metastasis.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

PD

XX

09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

31-JAN-2000; 2000US-0179065.

PR

04-FEB-2000; 2000US-0180628.

PR

24-FEB-2000; 2000US-0184664.

PR

02-MAR-2000; 2000US-0186350.

PR

16-MAR-2000; 2000US-0189874.

PR

17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216547.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226681.

PR 23-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0228824.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232080.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

CC AK54951 to AK64702 encode the human immune/hematopoietic antigen (1)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosol
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Claim 11: SEQ ID NO 18569; 3071pp + Sequence Listing; English.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX MPI: 2001-483426/52.
 DR N-PSDB; AAK63757.
 XX

Query Match	Best Local Similarity	1.9%; Score 8;	DB 22;	Length 126;
Matches	8;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;	
Qy	61 ASDYLLFT 68			
Db	20 ASDYLLFT 27			
RESULT 14				
AA079595	AA079595 standard; Protein: 128 AA.			
XX	AA079595;			
XX	06-NOV-2001 (first entry)			
DE	Human protein SEQ ID NO 3241.			
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;			
KW	vaccine; peptide therapy; stem cell growth factor; hematopoiesis;			
KW	tissue growth factor; immunomodulatory; cancer; leukemia;			
KW	nervous system disorder; arthritis; inflammation.			
XX	Homo sapiens.			
XX	WO200157190-A2.			
XX	09-AUG-2001.			
XX	05-FEB-2001; 2001WO-US04098.			
XX	03-FEB-2000; 2000US-046914.			
PR	27-APR-2000; 2000US-0560875.			
PR	20-JUN-2000; 2000US-0598075.			
PR	19-JUL-2000; 2000US-0620325.			
PR	01-SEP-2000; 2000US-0654936.			
PR	15-SEP-2000; 2000US-0663561.			
PR	20-OCT-2000; 2000US-0693325.			
PR	30-NOV-2000; 2000US-0728422.			
XX	(HYSE-) HYSEQ INC.			
PA	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;			
PI	Zhao XH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;			
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;			
XX	WPI; 2001-476283/51.			
DR	N-PSDB; AAK52728.			
XX	Nucleic acids encoding polypeptides with cytokine-like activities,			
XX	useful in diagnosis and gene therapy -			
PS	Claim 20; Page 293; 6221pp; English.			

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM8333-AAK0302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoietic regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM0020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 SQ Sequence 128 AA;
 Query Match 1.9%; Score 8; DB 22; Length 128;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDLITCLE 126
 |||||
 Db 101 PDLITCLE 108
 RESULT 15
 AAM40732
 ID AAM40732 standard; Protein; 193 AA.
 AC AAM40732:
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5663.
 XX
 KW Human: nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PE 26-DEC-2000; 2000MO-0534263.
 XX
 PR 21-JAN-2000; 2000US-048725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Iiu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI59888.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 5663; 10078bp; English.

XX
 CC the invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAK42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 193 AA;
 Query Match 1.9%; Score 8; DB 22; Length 193;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDLITCLE 126
 |||||
 Db 100 PDLITCLE 107
 Search completed: May 15, 2003, 13:22:39
 Job time : 45.6531 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:22:46 ; Search time 22.3108 Seconds
(without alignments)
1807.034 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDCGHLVTVYODSGTMT.....STRKDETTTRIPSEETST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 362588 seqs, 96450795 residues

Word size : 0

Total number of hits satisfying chosen parameters: 204442

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : Published Applications, AA.*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCY05_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	100.0	421	9	US-09-759-130B-76 Sequence 76, Appl
2	418	100.0	681	9	US-09-759-130B-75 Sequence 75, Appl
3	28	6.7	503	9	US-10-003-132-4 Sequence 4, Appl1
4	28	6.7	503	10	US-09-823-038A-51 Sequence 51, Appl
5	19	4.5	458	9	US-10-003-132-6 Sequence 6, Appl1
6	7	1.7	206	9	US-09-738-626-6397 Sequence 6397, Ap
7	7	1.7	258	9	US-09-738-626-5807 Sequence 5807, Ap
8	7	1.7	299	9	US-09-738-626-4484 Sequence 4484, Ap
9	7	1.7	383	10	US-09-909-849-2 Sequence 2, Appl1
10	7	1.7	385	9	US-09-712-363-148 Sequence 148, Appl
11	7	1.7	449	9	US-09-910-186A-14 Sequence 14, Appl
12	7	1.7	653	9	US-10-060-830-1114 Sequence 1114, Ap
13	6	1.4	100	10	US-09-864-761-47749 Sequence 47749, A
14	6	1.4	102	9	US-09-955-866-20 Sequence 20, Appl
15	6	1.4	108	9	US-09-949-510-1 Sequence 1, Appl1
16	6	1.4	108	10	US-09-910-150-21 Sequence 21, Appl
17	6	1.4	110	10	US-09-815-242-10553 Sequence 10553, A
18	6	1.4	111	10	US-09-925-300-1524 Sequence 1524, Ap
19	6	1.4	112	10	US-09-910-150-19 Sequence 19, Appl

20	6	1.4	116	10	US-09-800-729-183 Sequence 183, App
21	6	1.4	118	10	US-09-910-150-23 Sequence 23, Appl
22	6	1.4	119	10	US-09-867-550-410 Sequence 410, App
23	6	1.4	123	1	US-08-899-112-10 Sequence 10, Appl
24	6	1.4	123	8	US-08-424-508-64 Sequence 64, Appl
25	6	1.4	123	10	US-09-771-161A-171 Sequence 171, App
26	6	1.4	125	9	US-09-764-868-911 Sequence 911, App
27	6	1.4	135	10	US-09-925-301-1529 Sequence 1529, Ap
28	6	1.4	147	9	US-09-809-391-503 Sequence 503, App
29	6	1.4	147	10	US-09-864-761-48692 Sequence 48692, A
30	6	1.4	151	9	US-09-776-724A-262 Sequence 262, App
31	6	1.4	161	10	US-09-925-301-929 Sequence 929, App
32	6	1.4	175	9	US-10-108-605-97 Sequence 97, Appl
33	6	1.4	176	12	US-10-062-254-230 Sequence 130, App
34	6	1.4	178	10	US-09-737-178-132 Sequence 132, App
35	6	1.4	178	10	US-09-864-761-33694 Sequence 33694, A
36	6	1.4	178	10	US-09-864-761-46479 Sequence 46479, A
37	6	1.4	183	9	US-10-116-255-27 Sequence 27, Appl1
38	6	1.4	183	10	US-09-910-174A-4 Sequence 4, Appl1
39	6	1.4	191	9	US-10-260-877-56 Sequence 56, Appl
40	6	1.4	195	10	US-09-764-864-854 Sequence 854, App
41	6	1.4	201	9	US-09-924-340-56 Sequence 56, Appl
42	6	1.4	201	9	US-09-924-340-56 Sequence 56, Appl
43	6	1.4	201	10	US-09-924-340-56 Sequence 56, Appl
44	6	1.4	201	10	US-09-924-340-56 Sequence 56, Appl
45	6	1.4	202	10	US-09-924-340-56 Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-759-130B-76
Sequence 76, Application US/09759130B
Publication No. US2003002279A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: McCarthy, Sean A
APPLICANT: Fraser, Christopher C
APPLICANT: Sharp, John D
APPLICANT: Barnes, Thomas S
APPLICANT: Kirst, Susan J
APPLICANT: Mackay, Charles R
APPLICANT: Myers, Paul S
APPLICANT: Leiby, Kevin R
APPLICANT: Wrighton, Nicolas
APPLICANT: Goodearl, Andrew
APPLICANT: Holtzman, Douglas A
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE,
TITLE OF INVENTION: THERAPEUTIC, AND OTHER
FILE REFERENCE: MP100-5350ANIM
CURRENT APPLICATION NUMBER: US/09/759,130B
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: US 09/479,249
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: US 09/559,497
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 09/578,063
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: US 09/333,159
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: US 09/596,194
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/342,364
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: US 09/608,452
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/393,996
PRIOR FILING DATE: 1999-09-10
PRIOR APPLICATION NUMBER: US 09/602,871
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 09/420,707

```
;; PRIOR FILING DATE: 1999-10-19
;; NUMBER OF SEQ ID NOS: 460
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 76
;; LENGTH: 421
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-759-1308-76

Query Match
Best Local Similarity 100.0%; Score 418; DB 9; Length 421;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BELGDCGHLVYDDSGTMTSKNTPGYPNHTVCEKTTVPKGRLLRLGLDIESQTC 60
DB 1 BELGDCGHLVYDDSGTMTSKNTPGYPNHTVCEKTTVPKGRLLRLGLDIESQTC 60
QY 61 ASDYLFTSSDQYGPYCGSMTPPEKLLNTSEVTVRESGSHSGRGLTYASSDHPD 120
DB 61 ASDYLFTSSDQYGPYCGSMTPPEKLLNTSEVTVRESGSHSGRGLTYASSDHPD 120
QY 121 LITCLERASHLYKTEYSKFCPCAGCDVADISGNMVDGYRDTSLCKAIIHAGIADDELG 180
DB 121 LITCLERASHLYKTEYSKFCPCAGCDVADISGNMVDGYRDTSLCKAIIHAGIADDELG 180
QY 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRLFTSNGCSRSLSFEPDQIRASSSMQ 240
DB 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRLFTSNGCSRSLSFEPDQIRASSSMQ 240
QY 241 SVNESGDVHMSPGARLODQPSWASGDSNNHPRMLEIDLGKKKITGRTGSGTQ 300
DB 241 SVNESGDVHMSPGARLODQPSWASGDSNNHPRMLEIDLGKKKITGRTGSGTQ 300
QY 301 SNFNFYVKSFWANFNKNNSKMKTGYKGIYNNKRYVQGSNFRDPVQNNFIPPIVARYRV 360
DB 301 SNFNFYVKSFWANFNKNNSKMKTGYKGIYNNKRYVQGSNFRDPVQNNFIPPIVARYRV 360
QY 361 VPQTHORIALKVELIGQITQGNDSLWKRKTSOSTSVSTKKEDETTIRPIPSEETST 418
DB 361 VPQTHORIALKVELIGQITQGNDSLWKRKTSOSTSVSTKKEDETTIRPIPSEETST 418

RESULT 2
US-09-759-1308-75
;; Sequence 75, Application US/09759130B
;; Publication No. US20030022279A1
;; GENERAL INFORMATION:
;; APPLICANT: Millennium Pharmaceuticals, Inc.
;; APPLICANT: McCarthy, Sean A
;; APPLICANT: Fraser, Christopher C
;; APPLICANT: Sharp, John D
;; APPLICANT: Barnes, Thomas S
;; APPLICANT: Kirst, Susan J
;; APPLICANT: Mackay, Charles R
;; APPLICANT: Myers, Paul S
;; APPLICANT: Leiby, Kevin R
;; APPLICANT: Wrighton, Nicolas
;; APPLICANT: Goodearl, Andrew
;; APPLICANT: Holtzman, Douglas A
;; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
;; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
;; TITLE OF INVENTION: USES.
;; FILE REFERENCE: MP100-5350NM1M
;; CURRENT APPLICATION NUMBER: US/09/759,130B
;; CURRENT FILING DATE: 2002-09-16
;; PRIOR APPLICATION NUMBER: US 09/479,249
;; PRIOR FILING DATE: 2000-01-07
;; PRIOR APPLICATION NUMBER: US 09/559,497
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 09/578,063
;; PRIOR FILING DATE: 2000-05-24
;; PRIOR APPLICATION NUMBER: US 09/333,159
;; PRIOR FILING DATE: 1999-06-14
```

```
;; PRIOR APPLICATION NUMBER: US 09/596,194
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 09/342,364
;; PRIOR FILING DATE: 1999-06-29
;; PRIOR APPLICATION NUMBER: US 09/608,452
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/393,996
;; PRIOR FILING DATE: 1999-09-10
;; PRIOR APPLICATION NUMBER: US 09/602,871
;; PRIOR FILING DATE: 2000-06-23
;; PRIOR APPLICATION NUMBER: US 09/420,707
;; PRIOR FILING DATE: 1999-10-19
;; NUMBER OF SEQ ID NOS: 460
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 75
;; LENGTH: 681
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-759-1308-75

Query Match
Best Local Similarity 100.0%; Score 418; DB 9; Length 681;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BELGDCGHLVYDDSGTMTSKNTPGYPNHTVCEKTTVPKGRLLRLGLDIESQTC 60
DB 1 BELGDCGHLVYDDSGTMTSKNTPGYPNHTVCEKTTVPKGRLLRLGLDIESQTC 60
QY 61 ASDYLFTSSDQYGPYCGSMTPPEKLLNTSEVTVRESGSHSGRGLTYASSDHPD 120
DB 61 ASDYLFTSSDQYGPYCGSMTPPEKLLNTSEVTVRESGSHSGRGLTYASSDHPD 120
QY 121 LITCLERASHLYKTEYSKFCPCAGCDVADISGNMVDGYRDTSLCKAIIHAGIADDELG 180
DB 121 LITCLERASHLYKTEYSKFCPCAGCDVADISGNMVDGYRDTSLCKAIIHAGIADDELG 180
QY 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRLFTSNGCSRSLSFEPDQIRASSSMQ 240
DB 181 GOISVLQKRGISRYEGIIANGVLSRDGSLSDKRLFTSNGCSRSLSFEPDQIRASSSMQ 240
QY 241 SVNESGDVHMSPGARLODQPSWASGDSNNHPRMLEIDLGKKKITGRTGSGTQ 300
DB 241 SVNESGDVHMSPGARLODQPSWASGDSNNHPRMLEIDLGKKKITGRTGSGTQ 300
QY 301 SNFNFYVKSFWANFNKNNSKMKTGYKGIYNNKRYVQGSNFRDPVQNNFIPPIVARYRV 360
DB 301 SNFNFYVKSFWANFNKNNSKMKTGYKGIYNNKRYVQGSNFRDPVQNNFIPPIVARYRV 360
QY 361 VPQTHORIALKVELIGQITQGNDSLWKRKTSOSTSVSTKKEDETTIRPIPSEETST 418
DB 361 VPQTHORIALKVELIGQITQGNDSLWKRKTSOSTSVSTKKEDETTIRPIPSEETST 418

RESULT 3
US-10-003-132-4
;; Sequence 4, Application US/10003132
;; Publication No. US20020192750A1
;; GENERAL INFORMATION:
;; APPLICANT: Fox, Brian A.
;; APPLICANT: Gao, Zeren
;; APPLICANT: Shoemaker, Kimberly E.
;; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
;; FILE REFERENCE: 00-62
;; CURRENT APPLICATION NUMBER: US/10/003,132
;; CURRENT FILING DATE: 2001-11-15
;; PRIOR APPLICATION NUMBER: US 60/249,004
;; PRIOR FILING DATE: 2000-11-15
;; NUMBER OF SEQ ID NOS: 19
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 4
;; LENGTH: 503
;; TYPE: PRT
;; ORGANISM: Mus musculus
```

US-10-003-132-4

Query Match 6.7%; Score 28; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.2e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SGGHISGRGFLTYASSDHPDLITCER 127
DB 125 SGGHISGRGFLTYASSDHPDLITCER 152

RESULT 4

US-09-823-038A-51
; Sequence 51, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mouse
US-09-823-038A-51

Query Match 6.7%; Score 28; DB 10; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.2e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SGGHISGRGFLTYASSDHPDLITCER 127
DB 125 SGGHISGRGFLTYASSDHPDLITCER 152

RESULT 5

US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Fox, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUTROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6

Query Match 4.5%; Score 19; DB 9; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 DGYRDTSLCKAIIHAGII 175
DB 137 DGYRDTSLCKAIIHAGII 155

RESULT 6

US-09-738-626-6397
; Sequence 6397, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6397
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6397

Query Match 1.7%; Score 7; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 TYRFEFG 101
DB 75 TYRFEFG 81

RESULT 7

US-09-738-626-5807
; Sequence 5807, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5807
; LENGTH: 258

;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5807

Query Match 1.7%; Score 7; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 WASGDS 271
|||||
DB 240 WASGDS 246

RESULT 8
US-09-738-626-4484

;; Sequence 4484, Application US/09738626
;; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SEMOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738, 626
;; PRIOR FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 4484
;; LENGTH: 299
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4484

Query Match 1.7%; Score 7; DB 9; Length 299;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 TYPKELL 88
|||||
DB 16 TYPKELL 22

RESULT 9
US-09-909-849-2

;; Sequence 2, Application US/09909849
;; Patent No. US20020106754A1
;; GENERAL INFORMATION:
;; APPLICANT: Tauch, Andreas
;; TITLE OF INVENTION: Nucleotide Sequences Which Code for the a1r Gene
;; FILE REFERENCE: 032301 WD 173
;; CURRENT APPLICATION NUMBER: US/09/909, 849
;; PRIOR FILING DATE: 2001-07-23
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2
;; LENGTH: 383
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-909-849-2

Query Match 1.7%; Score 7; DB 10; Length 383;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 SWAGDS 270
|||||
DB 376 SWAGDS 382

RESULT 10
US-09-712-363-148

;; Sequence 148, Application US/09712363
;; Patent No. US20020164588A1
;; GENERAL INFORMATION:
;; APPLICANT: Eisenberg, David
;; APPLICANT: Rotstein, Sergio H.
;; APPLICANT: Marcotte, Edward M.
;; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;; FILE REFERENCE: 07419-032001
;; CURRENT APPLICATION NUMBER: US/09/712, 363
;; PRIOR FILING DATE: 2000-11-13
;; PRIOR APPLICATION NUMBER: PCT/US00/02246
;; PRIOR FILING DATE: 2000-01-28
;; PRIOR APPLICATION NUMBER: 60/179, 531
;; PRIOR FILING DATE: 2000-02-01
;; PRIOR APPLICATION NUMBER: 60/117, 844
;; PRIOR FILING DATE: 1999-01-29
;; PRIOR APPLICATION NUMBER: 60/118, 206,
;; PRIOR FILING DATE: 1999-02-01
;; PRIOR APPLICATION NUMBER: 60/126, 593
;; PRIOR FILING DATE: 1999-03-26
;; PRIOR APPLICATION NUMBER: 60/134, 093
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 60/134, 092
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 60/165, 124
;; PRIOR FILING DATE: 1999-11-12
;; PRIOR APPLICATION NUMBER: 60/165, 086
;; PRIOR FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 292
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 148
;; LENGTH: 385
;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-148

Query Match 1.7%; Score 7; DB 9; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 LILRLGD 52
|||||
DB 280 LILRLGD 286

RESULT 11
US-09-910-186A-14

;; Sequence 14, Application US/09910186A
;; Publication No. US20030009025A1
;; GENERAL INFORMATION:
;; APPLICANT: U.S. Army Medical Research & Materiel Command
;; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
;; FILE REFERENCE: A33626-A 067252.0107
;; CURRENT APPLICATION NUMBER: US/09/910, 186A
;; PRIOR FILING DATE: 2001-07-20
;; PRIOR APPLICATION NUMBER: PCT/US00/12890
;; PRIOR FILING DATE: 2000-05-12
;; PRIOR APPLICATION NUMBER: 09/611, 419
;; PRIOR FILING DATE: 2000-07-06
;; PRIOR APPLICATION NUMBER: 60/133, 865

;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,866
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,867
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,868
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,869
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 60/133,873
;; PRIOR FILING DATE: 1999-05-12
;; PRIOR APPLICATION NUMBER: 08/123,975
;; PRIOR FILING DATE: 1993-09-21
;; NUMBER OF SEQ ID NOS: 34
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 14
;; LENGTH: 449
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Synthetic Construct
US-09-910-186A-14

Query Match 1.7%; Score 7; DB 9; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 312 MNEKNN 318
|||||||
Db 396 MNEKNN 402

RESULT 12
US-10-060-830-1114
;; Sequence 1114, Application US/10060830
;; Publication No. US20030032154A1
;; GENERAL INFORMATION:
;; APPLICANT: Gu, Yizhong
;; APPLICANT: Nguyen, Cung-Tuong
;; TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
;; FILE REFERENCE: PB0169
;; CURRENT APPLICATION NUMBER: US/10/060,830
;; CURRENT FILING DATE: 2002-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/325,062
;; PRIOR FILING DATE: 2001-09-25
;; NUMBER OF SEQ ID NOS: 1123
;; SOFTWARE: Aeomica Sequence Listing Engine
;; SEQ ID NO 1114
;; LENGTH: 653
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-060-830-1114

Query Match 1.7%; Score 7; DB 9; Length 653;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 LGGQISV 185
|||||||

Db 128 LGGQISV 134

RESULT 13
US-09-864-761-47749
;; Sequence 47749, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;; FILE REFERENCE: Aeomica-x-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 47749
;; LENGTH: 100
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL109809.9
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.88
;; OTHER INFORMATION: EST HUMAN HIT: AW406955.1, EVALUATE 1.00e-01
;; OTHER INFORMATION: SWISSPROT HIT: P01871, EVALUATE 7.00e-03
US-09-864-761-47749

Query Match 1.4%; Score 6; DB 10; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 LTYASS 116
|||||||
Db 11 LTYASS 16

RESULT 14

US-09-955-866-20
; Sequence 20, Application US/09955866
; Patent No. US20020107363A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Michael
; APPLICANT: Sullivan, John K.
; APPLICANT: Holst, Paige
; APPLICANT: Yoshinaga, Steven Kiyoshi
; TITLE OF INVENTION: B7-Like Polypeptides and Uses Thereof
; FILE REFERENCE: 00,759-A
; CURRENT APPLICATION NUMBER: US/09/955,866
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/233,867
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 20
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-866-20

Query Match

1.43; Score 6; DB 10; Length 102;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 TPKEL 87
|||||
Db 6 TPKEL 11

RESULT 15

US-09-949-510-1
; Sequence 1, Application US/09949510
; Publication No. US20030077840A1
; GENERAL INFORMATION:
; APPLICANT: Chait et al.
; TITLE OF INVENTION: METHOD FOR THE COMPARATIVE QUANTITATIVE ANALYSIS OF
; TITLE OF INVENTION: PROTEINS AND OTHER BIOLOGICAL MATERIAL BY ISOTOPIC
; TITLE OF INVENTION: LABELING AND MASS SPECTROSCOPY
; FILE REFERENCE: Seq. List 1-6
; CURRENT APPLICATION NUMBER: US/09/949,510
; CURRENT FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Human
US-09-949-510-1

Query Match

1.43; Score 6; DB 9; Length 108;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 ISRYEG 196
|||||

Db 47 ISRYEG 52

Search completed: May 15, 2003, 13:26:35
Job time : 25.3108 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:21:41 ; Search time 36.7473 Seconds
(without alignments)
1093.529 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDGCGHLVYDSCGTMF.....STKKDETTTRIPSEST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 223049

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.2	440	1	RGECUC
2	9	2.2	440	2	D91204
3	9	2.2	440	2	F86050
4	8	1.9	154	2	D39384
5	8	1.9	163	2	B39384
6	8	1.9	195	2	E39384
7	8	1.9	196	2	A39384
8	8	1.9	255	2	B97861
9	8	1.9	264	2	AB2326
10	8	1.9	595	2	G02075
11	8	1.9	624	2	S50650
12	7	1.7	100	2	A82629
13	7	1.7	117	2	T49511
14	7	1.7	167	2	T33602
15	7	1.7	182	2	T34009
16	7	1.7	183	2	D84183
17	7	1.7	213	2	E96943
18	7	1.7	248	2	D69713
19	7	1.7	254	2	T23323
20	7	1.7	260	2	F85630
21	7	1.7	261	2	C83756
22	7	1.7	265	2	S20988
23	7	1.7	272	2	B90544
24	7	1.7	273	2	T03254
25	7	1.7	285	2	S27365
26	7	1.7	285	2	G83934
27	7	1.7	293	2	G64050
28	7	1.7	301	2	T19399
29	7	1.7	306	2	D70601

30	7	1.7	306	2	T45453	UTP-glucose-1-phos
31	7	1.7	322	2	AC1759	conserved hypothet
32	7	1.7	332	2	A11383	conserved hypothet
33	7	1.7	339	2	T37727	probable nuclear p
34	7	1.7	356	1	S00139	photosynthetic rea
35	7	1.7	357	2	C69223	anion permease - M
36	7	1.7	359	2	T21373	hypothetical prote
37	7	1.7	385	2	S70984	recf protein - Myc
38	7	1.7	394	2	A70842	probable antib prot
39	7	1.7	397	2	F85681	unknown protein en
40	7	1.7	399	2	T01035	hypothetical prote
41	7	1.7	400	2	F81419	probable efflux pr
42	7	1.7	405	2	C71462	hypothetical prote
43	7	1.7	414	2	F29826	hypothetical prote
44	7	1.7	444	2	S34200	polygalacturonase
45	7	1.7	445	2	S34266	polygalacturonase

ALIGNMENTS

RESULT 1

RGECUC Regulatory protein uhpc - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 30-Jun-1991 #sequence_revision 14-Nov-1997 #extl_change 01-Mar-2002

C:Accession: D65168; G41853; C26925; S30078

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; M01D:97426617; PMID:9278503

A:Accession: D65168

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-440 <BLAT>

A:Cross-references: GB:AE000444; GB:U00096; NID:q2367258; PIDN:AA24726.1; PID:q148119

A:Experimental source: strain K-12, substrain MG1655

J. Bacteriol. 174, 2754-2762, 1992

A:Title: Structure and function of the uhpc genes for the sugar phosphate transport sy

A:Reference number: A41853; M01D:9224930; PMID:1569007

A:Accession: G41853

A:Molecule type: DNA

A:Residues: 2-440 <RSL>

A:Cross-references: GB:M89479; NID:q148116; PIDN:AA24726.1; PID:q148119

A:Note: this is a revision to the sequence from reference A30395

R:Frledrich, M.J.; Kadner, R.J.

J. Bacteriol. 169, 3556-3563, 1987

A:Title: Nucleotide sequence of the uhpc region of Escherichia coli.

A:Reference number: A30395; M01D:87279903; PMID:3301805

A:Accession: C26925

A:Molecule type: DNA

A:Residues: 'M', '87', 'VCRAIAE', '96', 'L', '98', 'P', '100', 'L', '102', 'FPG', '139-141', 'SVNGLVFTYRA', '1

A:Note: this sequence has been revised in reference A41853

C:Comment: This is one of the proteins involved in the expression of uhpc, a gene for

NAD transmembrane orientation.

C:Genetics:

A:Gene: uhpc

A:Map position: 82 min

C:Superfamily: hexose phosphate transport protein uhpc

C:Keywords: membrane protein; sugar phosphate transport system

Query Match Best Local Similarity 2.2% Score 9: DB 1: Length 440:

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 197 ILANGVLSR 205

Db 57 ILANGVLSR 65

RESULT 2

D91204
 regulator of uhpt [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: D91204
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A96629; MUID:21156231; PMID:11258796
 A:Accession: D91204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-440 <HAY>
 A:Cross-references: GB:BA000007; PID:BA838027.1; PID:Q13364079; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: EC64604
 C:Superfamily: hexose phosphate transport protein uhpt

Query Match 2.2%; Score 9; DB 2; Length 440;
 Best Local Similarity 100.0%; Pred. No. 0.62;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
 |||||||||
 DB 57 ILANGVLSR 65

RESULT 3

regulator of uhpt [imported] - Escherichia coli (strain O157:H7, substrain EDJ933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: F86050
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: F86050
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-440 <STO>
 A:Cross-references: GB:AE005174; NID:Q12518499; PID:AA658866.1; GSPDB:GN00145; UWGP:251
 A:Experimental source: strain O157:H7, substrain EDJ933
 C:Genetics:
 A:Gene: uhpc
 C:Superfamily: hexose phosphate transport protein uhpt

Query Match 2.2%; Score 9; DB 2; Length 440;
 Best Local Similarity 100.0%; Pred. No. 0.62;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
 |||||||||
 DB 57 ILANGVLSR 65

RESULT 4

finger protein HRF6 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
 C:Accession: D39384
 R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
 A:Title: The evolutionarily conserved Kruempel-associated box domain defines a subfamily
 A:Reference number: A39384; MUID:91219421; PMID:2023909
 A:Accession: D39384
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <BEL>

A:Cross-references: GB:M61869; NID:Q184449; PID:Q184450
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||||
 DB 8 PDLITCLE 15

RESULT 5

finger protein HRP9 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 08-Dec-2000
 C:Accession: B39384
 R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
 A:Title: The evolutionarily conserved Kruempel-associated box domain defines a subfam
 A:Reference number: A39384; MUID:91219421; PMID:2023909
 A:Accession: B39384
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-163 <AAA>
 A:Cross-references: GB:M61867
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 163;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||||
 DB 18 PDLITCLE 25

RESULT 6

finger protein HRF9 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 20-Mar-1992 #sequence_revision 30-Jan-1993 #text_change 08-Dec-2000
 C:Accession: E39384
 R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
 A:Title: The evolutionarily conserved Kruempel-associated box domain defines a subfam
 A:Reference number: A39384; MUID:91219421; PMID:2023909
 A:Accession: E39384
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-195 <BEL>
 A:Cross-references: GB:M61869
 C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 195;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||||
 DB 50 PDLITCLE 57

RESULT 7

finger protein HRF4 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
 C:Accession: A39384
 R:Bellefroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Krueppel-associated box domain defines a subfamily
A:Reference number: A39384; MID:91219421; PMID:2023909
A:Accession: A39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <REL>
A:Cross-references: GB:M61866; MID:9454818; PIDN:AAA52689.1; PID:g184336
C:Keywords: DNA binding; zinc finger

Query Match 1.9%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCIE 126
|||||||
DB 50 PDLITCIE 57

RESULT 8
B97861
hypothetical protein RC1290 [Imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: B97861
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MID:21442074; PMID:11557893
A:Accession: B97861
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <RUR>
A:Cross-references: GB:AE006914; PIDN:AAL03828.1; PID:g15620429; GSPDB:GN00173
C:Genetics:
A:Gene: RC1290

Query Match 1.9%; Score 8; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 ALKVELIG 377
|||||||
DB 242 ALKVELIG 249

RESULT 9
AB2326
hypothetical protein alr4161 [Imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AB2326
R:Kranek, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MID:21595285; PMID:11759840
A:Accession: AB2326
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <RUR>
A:Cross-references: GB:BA000019; PIDN:BA075860.1; PID:g17133296; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr4161

Query Match 1.9%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 357 YVRVPOQT 364
|||||||

DB 250 YVRVPOQT 257

RESULT 10
G02075
transcription repressor zinc finger protein 85 - human
C:Species: Homo sapiens (hmn)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
C:Accession: G02075
R:Poncellet, D.A.
submitted to the EMBL Data Library, September 1995
A:Reference number: G09169
A:Accession: G02075
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-595 <PON>
A:Cross-references: EMBL:U5376; MID:g1017721; PIDN:AAA9179.1; PID:g1017722
C:Genetics:
A:Gene: GDB:ZNF85
A:Cross-references: GDB:132279
A:Map position: 19p12-19p12
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 1.9%; Score 8; DB 2; Length 595;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCIE 126
|||||||
DB 50 PDLITCIE 57

RESULT 11
S50650
hypothetical protein YER147c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
C:Accession: S50650
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and 1amb
A:Reference number: S50430
A:Accession: S50650
A:Molecule type: DNA
A:Residues: 1-624 <DIE>
A:Cross-references: EMBL:U18917; MID:g603377; PID:g603387; GSPDB:GN00005; MIPS:YER147
C:Genetics:
A:Gene: SGD:SCC4; MIPS:YER147c
A:Cross-references: SGD:S0000949
A:Map position: 5R
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YER147c

Query Match 1.9%; Score 8; DB 2; Length 624;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 368 RIALKVEL 375
|||||||
DB 285 RIALKVEL 292

RESULT 12
AB2629
hypothetical protein XF1865 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: AB2629
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: AB2629

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-100 <SIR>
 A:Cross-references: GB:AE004007; GB:AE003849; NID:g9106944; PIDN:AFR64671.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reimach, F.C.; Arruda, P.; Abreu, F.A.; Agencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincan, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XFI865

Query Match 1.7%; Score 7; DB 2; Length 100;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 50 LGDDIE 56
 |||||
 Db 55 LGDDIE 61

RESULT 13
 T49511
 hypothetical protein B14D6.630 [Imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
 C:Accession: T49511
 R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25022
 A:Accession: T49511
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-117 <SCH>
 A:Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.630
 A:Experimental source: BAC clone B14D6; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B14D6.630
 A:Map position: 6
 C:Superfamily: Neurospora crassa hypothetical protein B14D6.630

Query Match 1.7%; Score 7; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 GHLVYQ 14
 |||||
 Db 4 GHLVYQ 10

RESULT 14
 T33602
 hypothetical protein E02H9.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33602
 R:Kellen, J.; Kramer, J.; Hawkins, M.
 submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of C. elegans cosmid E02H9.
 A:Reference number: Z21375
 A:Accession: T33602
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA

A:Residues: 1-167 <KEI>
 A:Cross-references: EMBL:AF099915; PIDN:AAC68768.1; GSPDB:GN00021; CESP:E02H9.2
 A:Experimental source: strain Bristol N2; clone E02H9
 C:Genetics:
 A:Gene: CESP:E02H9.2
 A:Map position: 3
 A:Insertions: 89/2

Query Match 1.7%; Score 7; DB 2; Length 167;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 RKGISRY 194
 |||||
 Db 55 RKGISRY 61

RESULT 15
 T34009
 hypothetical protein Y49G5A.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34009
 R:Becker, M.; Tin-Wollam, A.M.; Yeakum, M.
 submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of C. elegans cosmid Y49G5A.
 A:Reference number: Z21457
 A:Accession: T34009
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-182 <BEC>
 A:Cross-references: EMBL:AF125968; PIDN:AAD14760.1; GSPDB:GN00023; CESP:Y49G5A.1
 A:Experimental source: strain Bristol N2; clone Y49G5A
 C:Genetics:
 A:Gene: CESP:Y49G5A.1
 A:Map position: 5
 A:Insertions: 44/3; 82/1; 102/3

Query Match 1.7%; Score 7; DB 2; Length 182;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 SKFCPAG 143
 |||||
 Db 162 SKFCPAG 168

Search completed: May 15, 2003, 13:25:55
 Job time : 39.7473 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:20:51 ; Search time 24.9356 Seconds
(without alignments)
493.221 Million cell updates/sec.

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EELGDCGHLVYQDSGTWT.....STKKEDEITRPISERNST 418

Scoring table: OLIGO
Gapco 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 58762

Minimum DB seq length: 100
Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgcn2_6/ptodata/1/1aa/5A.COMB.pep:*\n2: /cgcn2_6/ptodata/1/1aa/5B.COMB.pep:*\n3: /cgcn2_6/ptodata/1/1aa/5A.COMB.pep:*\n4: /cgcn2_6/ptodata/1/1aa/5B.COMB.pep:*\n5: /cgcn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*\n6: /cgcn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	1.7	115	1	US-08-111-939-22
2	7	1.7	159	2	US-08-162-402B-16
3	7	1.7	218	1	US-07-607-538C-4
4	7	1.7	218	2	US-08-162-402B-4
5	7	1.7	272	2	US-08-492-027A-4
6	6	1.4	108	4	US-09-304-799-1
7	6	1.4	109	1	US-08-111-939-21
8	6	1.4	109	1	US-08-111-939-25
9	6	1.4	111	1	US-08-111-939-23
10	6	1.4	115	4	US-09-134-001C-4964
11	6	1.4	123	2	US-08-626-685A-10
12	6	1.4	123	4	US-08-469-260A-64
13	6	1.4	146	1	US-08-054-480-2
14	6	1.4	147	4	US-09-149-476-503
15	6	1.4	151	2	US-08-563-148E-6
16	6	1.4	151	4	US-09-194-139-1
17	6	1.4	157	2	US-08-162-402B-13
18	6	1.4	160	2	US-08-162-402B-14
19	6	1.4	160	2	US-08-162-402B-15
20	6	1.4	176	4	US-09-393-627B-17
21	6	1.4	176	4	US-09-393-627B-18
22	6	1.4	188	2	US-08-531-525-46
23	6	1.4	188	2	US-08-718-270A-46
24	6	1.4	201	2	US-08-726-228-2
25	6	1.4	201	3	US-08-870-815-2
26	6	1.4	201	4	US-08-949-004-2
27	6	1.4	202	1	US-08-274-318-2

28	6	1.4	202	2	US-08-463-081B-2	Sequence 2, Appl1
29	6	1.4	202	2	US-08-461-379A-2	Sequence 2, Appl1
30	6	1.4	202	2	US-08-462-390B-2	Sequence 2, Appl1
31	6	1.4	202	2	US-08-754-108-2	Sequence 2, Appl1
32	6	1.4	202	3	US-08-870-815-4	Sequence 2, Appl1
33	6	1.4	202	3	US-08-463-074B-2	Sequence 4, Appl1
34	6	1.4	202	3	US-08-465-585C-2	Sequence 2, Appl1
35	6	1.4	202	3	US-08-652-446-2	Sequence 2, Appl1
36	6	1.4	202	4	US-08-949-004-4	Sequence 4, Appl1
37	6	1.4	208	4	US-08-961-083-214	Sequence 214, App
38	6	1.4	209	2	US-08-248-839C-58	Sequence 58, Appl
39	6	1.4	214	4	US-08-861-774E-48	Sequence 48, Appl
40	6	1.4	214	4	US-07-607-538C-5	Sequence 5, Appl1
41	6	1.4	218	2	US-08-162-402B-5	Sequence 5, Appl1
42	6	1.4	219	4	US-08-934-627B-6	Sequence 6, Appl1
43	6	1.4	221	4	US-09-198-956-4	Sequence 4, Appl1
44	6	1.4	221	4	US-09-670-141-4	Sequence 4, Appl1
45	6	1.4	227	4	US-08-871-572B-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-111-939-22
; Sequence 22, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
; APPLICANT: Kawai, Shinji
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Amano, Egon
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
; TITLE OF INVENTION: Protein and Process for its Production
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/111,939
; FILING DATE: 26-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324033/92
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 230029/92
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1321-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4000
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-111-939-22

Query Match 1.7%; Score 7; DB 1; Length 115;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 351 PRIVARY 357
Db 88 PRIVARY 94

RESULT 2
US-08-162-402B-16
Sequence 16, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KALTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-16

Query Match 1.7%; Score 7; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 351 PRIVARY 357
Db 129 PRIVARY 135

RESULT 3
US-07-607-538C-4
Sequence 4, Application US/07607538C
Patent No. 5455031
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KALTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

TITLE OF INVENTION: POLYPEPTIDE WITH 46
TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
TITLE OF INVENTION: USE THEREOF
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: V. Amzel & Assoc.
STREET: 2055 No. 5455031th Broadway
CITY: Walnut Creek
STATE: California
COUNTRY: USA
ZIP: 94596
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: CRFC-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-4

Query Match 1.7%; Score 7; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 351 PRIVARY 357
Db 28 PRIVARY 34

RESULT 4
US-08-162-402B-4
Sequence 4, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KALTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, VIVIANA
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-4

Query Match 1.7%; Score 7; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 PRIVATE 357
|||||
DB 28 PRIVATE 34

RESULT 5
US-08-492-027A-4
Sequence 4, Application US/08492027A
Patent No. 591233
GENERAL INFORMATION:
APPLICANT: Suzuki, Sholchi
APPLICANT: Burnell, James N
TITLE OF INVENTION: DNA Encoding Carbonic Anhydrase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/492,027A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0760-206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-492-027A-4

Query Match 1.7%; Score 7; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ALKVELI 376
|||||
DB 162 ALKVELI 168

RESULT 6
US-09-304-799-1
Sequence 1, Application US/09304799
Patent No. 6391649
GENERAL INFORMATION:
APPLICANT: The Rockefeller University
TITLE OF INVENTION: METHOD FOR THE COMPARATIVE ANALYSIS OF PROTEINS AND OTHER BIOL
FILE REFERENCE: 1119-0002
CURRENT APPLICATION NUMBER: US/09/304,799
CURRENT FILING DATE: 1999-05-04
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Version 3.0
SEQ ID NO 1
LENGTH: 108
TYPE: PRT
ORGANISM: Homo sapiens
US-09-304-799-1

Query Match 1.4%; Score 6; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 2,6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 ISRYEG 196
|||||
DB 47 ISRYEG 52

RESULT 7
US-08-111-939-21
Sequence 21, Application US/08111939
Patent No. 5460951
GENERAL INFORMATION:
APPLICANT: Kawai, Shinji
APPLICANT: Takeshita, Sunao
APPLICANT: Okazaki, Makoto
APPLICANT: Amann, Egon
TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
TITLE OF INVENTION: Protein and Process for its Production
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/111,939
FILING DATE: 26-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 324033/92
FILING DATE: 03-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 230029/92
FILING DATE: 28-AUG-1992

ATTORNEY/AGENT INFORMATION:
NAME: Forman, David S.
REGISTRATION NUMBER: 33,694
REFERENCE/DOCKET NUMBER: 02481.1321-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-111-939-21

Query Match 1.4%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 KVFQGN 338
|||||
DB 64 KVFQGN 69

RESULT 8
US-08-111-939-25
Sequence 25, Application US/08111939
Patent No. 5460951
GENERAL INFORMATION:
APPLICANT: Kawai, Shinji
APPLICANT: Takeshita, Sunao
APPLICANT: Okazaki, Makoto
APPLICANT: Amann, Egon
TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
NUMBER OF SEQUENCES: 27
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/111,939
FILING DATE: 26-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 324033/92
FILING DATE: 03-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 230029/92
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Forman, David S.
REGISTRATION NUMBER: 33,694
REFERENCE/DOCKET NUMBER: 02481.1321-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-111-939-25

Query Match 1.4%; Score 6; DB 1; Length 109;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 KVFQGN 338
|||||
DB 64 KVFQGN 69

RESULT 9
US-08-111-939-23
Sequence 23, Application US/08111939
Patent No. 5460951
GENERAL INFORMATION:
APPLICANT: Kawai, Shinji
APPLICANT: Takeshita, Sunao
APPLICANT: Okazaki, Makoto
APPLICANT: Amann, Egon
TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
NUMBER OF SEQUENCES: 27
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/111,939
FILING DATE: 26-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 324033/92
FILING DATE: 03-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 230029/92
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Forman, David S.
REGISTRATION NUMBER: 33,694
REFERENCE/DOCKET NUMBER: 02481.1321-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 111 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-111-939-23

Query Match 1.4%; Score 6; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 WLEIDL 284
|||||
DB 12 WLEIDL 17

RESULT 10
US-09-134-001C-4964
Sequence 4964, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4964
LENGTH: 115
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4964

Query Match 1.4%; Score 6; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 DYLLFT 68
DB 73 DYLLFT 78

RESULT 11
US-08-626-685A-10
Sequence 10, Application US/0862685A
Patent No. 5972624
GENERAL INFORMATION:
APPLICANT: Smith, Kelli E.
APPLICANT: Gerald, Christophe P.G.
APPLICANT: Weinschenk, Richard L.
APPLICANT: Linemeyer, David
APPLICANT: Branche, Theresa
APPLICANT: Forray, Carlos
TITLE OF INVENTION: DNA ENCODING A MAMMALIAN
TITLE OF INVENTION: HYPOTHALAMIC GALANIN RECEPTOR
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/626,685A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 50233-B/JPW/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-626-685A-10

Query Match 1.4%; Score 6; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 SLICKA 168
DB 95 SLICKA 100

RESULT 12
US-08-469-260A-64
Sequence 64, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAM J. PILOT-MARTIAS
APPLICANT: GEORGE J. DAWSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MURROFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIK
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: POREBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-260A-64

Query Match 1.4%; Score 6; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 ASSSMQ 240
DB 22 ASSSMQ 27

RESULT 13
US-08-054-480-2

Sequence 2, Application US/08054480
Patent No. 5525504
GENERAL INFORMATION:
APPLICANT: Goebel, Werner
APPLICANT: Libby, Stephen
APPLICANT: Heffton, Fred
TITLE OF INVENTION: CYTOLYSIN GENE AND GENE PRODUCT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLEN, WHITE, ZELANO, & BRANIGAN, P.C.
STREET: 2200 CLARENDON BOULEVARD, SUITE 1400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/054,480
FILING DATE: 04-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heaney, Brian P.
REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: MERCK 1496
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703 243 6333
TELEFAX: 703 243 6410
TELEX: 64191
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: SALMONELLA
US-08-054-480-2

Query Match 1.4%; Score 6; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 QTCASD 63
| | | | |
DB 81 QTCASD 86

RESULT 14
US-09-149-476-503
Sequence 503, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
EARLIER FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07

EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315

```

EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11

```

```

EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

```

```

Query Match 1.4%; Score 6; DB 4; Length 147;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 245 SGOVH 250
DB 62 SGOVH 67

```

```

RESULT 15
US-08-563-148E-6
Sequence 6, Application US/08563148E
Patent No. 5929224
GENERAL INFORMATION:
APPLICANT: Kazuo SUZUKI et al.
TITLE OF INVENTION: NOVEL ACTIVATING FACTOR OF LEUKOCYTES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: 663-2, Shitagi
STREET: Misaki-machi
CITY: Isumi-gun
STATE: Chiba-ken
COUNTRY: JAPAN
ZIP: 299-45
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS v.5
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/563,148E
FILING DATE: No. 5929224ember 27, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 293233/1994
FILING DATE: 28-NOV-1994
ATTORNEY/AGENT INFORMATION:

```

NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single strand
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: intermediate fragment
ORIGINAL SOURCE:
ORGANISM: human
TISSUE TYPE: T-cell derived leukemia cells
PUBLICATION INFORMATION:
AUTHORS: Kazuo SUZUKI et al.
TITLE: NOVEL ACTIVATING FACTOR OF LEUKOCYTES
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 to 151
US-08-563-148E-6

Query Match 1.43; Score 6; DB 2; length 151;
Best Local Similarity 100.08; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 ISGRGF 109
|||||
Db 93 ISGRGF 98

Search completed: May 15, 2003, 13:24:50
Job time : 26.9356 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:17:05 ; Search time 12.4678 Seconds

(Without alignments)
1390.548 Million cell updates/sec

Title: US-10-003-132-2_COPY_35_452

Perfect score: 418
Sequence: 1 EEELGDCGCHLVYODSGTMT.....STKDEDEITRPIPESETST 418

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 89080

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.2	439	1 UHPC_ECOLI	P09836 escherichia
2	8	1.9	195	1 ZN90_HUMAN	Q03938 homo sapien
3	8	1.9	353	1 Z234_HUMAN	Q75437 homo sapien
4	8	1.9	418	1 Z117_HUMAN	Q03924 homo sapien
5	8	1.9	535	1 Z257_HUMAN	Q9Y261 homo sapien
6	8	1.9	574	1 YE73_HUMAN	Q9P255 homo sapien
7	8	1.9	595	1 ZN85_HUMAN	Q03923 homo sapien
8	8	1.9	624	1 YEX7_YEAST	P40090 saccharomyc
9	7	1.7	248	1 S2SA_BACSU	Q34853 bacillus su
10	7	1.7	265	1 LECN_PEA	P16270 pisum sativ
11	7	1.7	280	1 XLRL_FUGRU	Q9W675 fugu rubrip
12	7	1.7	285	1 LEC_SOVBN	P05046 glycine max
13	7	1.7	293	1 NP40_HAEN	P44539 haemophilus
14	7	1.7	301	1 OM40_CAEEL	Q18090 caenorhabdi
15	7	1.7	322	1 Y073_LISMO	P58588 listeria in
16	7	1.7	322	1 Y016_LISIN	Q928C0 listeria in
17	7	1.7	339	1 SEHL_SCHPO	Q10099 schizosacch
18	7	1.7	356	1 CYCR_RHOVI	P07173 rhodospendo
19	7	1.7	359	1 MCAL_CRIGR	O54873 cricetus
20	7	1.7	385	1 REGT_MYCTU	O59386 mycobacteri
21	7	1.7	444	1 GEL2_ARATH	P49063 arabidopsis
22	7	1.7	445	1 ASSY_PASMU	P57877 pasteurella
23	7	1.7	446	1 ASSY_ECO57	O8X5M0 escherichia
24	7	1.7	446	1 ASSY_ECOLI	P22767 escherichia
25	7	1.7	446	1 PSES_PSEAE	O04004 pseudomonas
26	7	1.7	484	1 ER60_SCHMA	P38658 schistosoma
27	7	1.7	514	1 G6P1_MOUSE	Q00612 mus musculu
28	7	1.7	514	1 G6PD_MACRO	Q29492 macropus ro
29	7	1.7	514	1 G6PD_RAT	P05370 rattus norv
30	7	1.7	642	1 YM93_YEAST	P40308 saccharomyc
31	7	1.7	680	1 SYG_BOMMO	O04451 bombyx mori
32	6	1.4	102	1 RPOL_ARPE	Q9Y621 aeropyrum p
33	6	1.4	109	1 RLAI_SCHPO	P17476 schizosacch

34	6	1.4	109	1 RLAI_SCHPO	Q9uu78 schizosacch
35	6	1.4	110	1 KDFG_ERWCH	O05527 erwina chr
36	6	1.4	110	1 RLAI_SCHPO	P17477 schizosacch
37	6	1.4	114	1 MSME_MACMU	P25142 macaca mula
38	6	1.4	114	1 YG4U_YEAST	P53308 saccharomyc
39	6	1.4	125	1 HPAA_HELIAC	Q47947 helicobacte
40	6	1.4	129	1 CY52_SCHPO	Q9uam6 schizosacch
41	6	1.4	130	1 CCRN_RANCA	P80344 rana catesb
42	6	1.4	130	1 RS8_MERTG	Q977V0 methanococc
43	6	1.4	130	1 RS8_MERTJ	P54041 methanococc
44	6	1.4	130	1 RS8_MERTL	Q977U8 methanococc
45	6	1.4	133	1 PEMK_ECOLI	P13976 escherichia

ALIGNMENTS

RESULT 1
UHPC_ECOLI STANDARD; PRT; 439 AA.
AC P09836: P76728:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Regulatory protein uhpc.
GN UHPC OR B3667.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87279903; PubMed=3301805;
RA Friedrich M.J., Kadner R.J.;
RT "Nucleotide sequence of the uhpc region of Escherichia coli.";
RL J. Bacteriol. 169:3556-3563(1987).
RN [2]
RP REVISIONS.
RA Kadner R.J.;
RT Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92234930; PubMed=1569007;
RA Ireland M.D., Wei B.-Y., Kadner R.J.;
RT "Structure and function of the uhpc genes for the sugar phosphate transport system in Escherichia coli and Salmonella typhimurium.";
RL J. Bacteriol. 174:2754-2762(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=93315143; PubMed=7686882;
RT Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
RN [5]
RP POSTIBLE ROLE.
RA MEDLINE=93352407; PubMed=8349544;
RT Island M.D., Kadner R.J.;
RT "Interplay between the membrane-associated uhpc and uhpc regulatory proteins.";
RL J. Bacteriol. 175:5028-5034(1993).
RT J. FUNCTION: UHPC IS REQUIRED FOR UHPT EXPRESSION, IT MAY ACT JOINTLY WITH THE UHPC SIGNALING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@lsb-sib.ch.

CC EMBL; M17102; AAA24722.1; ALT_INIT.
CC EMBL; M89479; AAA24726.1; -
DR EMBL; L10328; AAA62019.1; ALT_INIT.
DR EMBL; AE000444; AAC76690.1; ALT_INIT.
DR PIR; G41853; RGEUC.
DR Ecocore; EG11053; unpc.
DR InterPro; IPR000849; G1PT_transporter.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr.1.
DR TIGRfams; TIGR00881; 2A0104; 1.
DR PROSITE; PS00942; G1PT.1.
KW Transport; Sugar transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 290 310 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 350 370 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
SO SEQUENCE 439 AA; 48256 MW; D8097864B77700 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 ILANGVLSR 205
Db 56 ILANGVLSR 64
|||||
ID ZN90_HUMAN STANDARD; PRT; 195 AA.
AC 003938;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 90 (Zinc finger protein HTR9) (Fragment).
GN ZNF90.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed-2023909;
RX MEDLINE-91219421; Poncelet D.A., Lecocq P.J., Revelant O.,
RA Bellefroid E.J., Poncelet D.A., Martiel J.A.;
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; M61870; AAA36028.1; -

PIR; E39384; E39384.
DR Genew; HGNC:13165; ZNF90.
DR MIM; 603973; -
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; ZF-C2H2; 1.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZNF_C2H2; 1.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 4 75 KRAB.
FT ZN_FING 145 167 C2H2-TYPE (DEGENERATE).
FT ZN_FING 173 195 C2H2-TYPE.
FT NON_TER 195 195
SO SEQUENCE 195 AA; 22562 MW; 0B808A1A62342608 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
Db 50 PDLITCLE 57
|||||
ID Z254_HUMAN STANDARD; PRT; 353 AA.
AC 075437;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 254 (Bone marrow zinc finger 5) (BMZF-5)
DE (Hematopoietic cell derived zinc finger protein 1) (HD-ZNF1).
GN ZNF254 OR BMZF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Blood;
RX MEDLINE-98318631; PubMed-9653160;
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
RA Wang Y.-X., Chen S.-J., Chen Z.;
RT "Identification of genes expressed in human CD34(+) hematopoietic
RT stem/progenitor cells by expressed sequence tags and efficient full-
RT length cDNA cloning";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
RN [2]
RP SEQUENCE FROM N.A. PubMed-10585455;
RX MEDLINE-20034457;
RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
RT hematopoietic cells and identification of a novel transregulatory
RT domain KRAB";
RL J. Biol. Chem. 274:35741-35748(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF054180; AAC39913.1; -

DR Genew: HGNC:13047; ZNF254.

DR MIM: 604768; -

DR InterPro: IPR001909; KRAB.

DR InterPro: IPR000822; Znf_C2H2.

DR Pfam: PF00096; zf-C2H2; 4.

DR Pfam: PF01352; KRAB; 1.

DR PRINTS: PRO00048; ZINC_FINGER.

DR PRODOM: PD000003; Znf_C2H2; 2.

DR SMART: SM00349; KRAB; 1.

DR SMART: SM00355; Znf_C2H2; 3.

DR PROSITE: PS50805; KRAB; 1.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.

DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 4.

DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;

KW Nuclear protein; Repeat.

FT DOMAIN 13 84 KRAB.

FT ZN_FING 210 232 ZINC_FINGERS.

FT ZN_FING 210 232 C2H2-TYPE (DEGENERATE).

FT ZN_FING 238 260 C2H2-TYPE (DEGENERATE).

FT ZN_FING 266 288 C2H2-TYPE.

FT ZN_FING 294 316 C2H2-TYPE.

FT ZN_FING 316 316 C2H2-TYPE.

SO SEQUENCE 353 AA; 41698 MW; D923CB82D2C5B56A CRC64;

Query Match 1.9%; Score 8; DB 1; Length 353;

Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126

DB 59 PDLITCLE 66

RESULT 4

ID 2117_HUMAN STANDARD; PRT; 418 AA.

AC 003924;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Zinc finger protein 117 (Zinc finger protein HPP9) (Fragment).

GN ZNF117.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE OF 4-418 FROM N.A.

RC TISSUE-Placenta;

RX MEDLINE-89377476; PubMed-2505992;

RA Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,

RA Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,

RT "The human genome contains hundreds of genes coding for finger

RT proteins of the Kruppel type.";

RL DNA 8:377-387(1989).

RN (2)

RP SEQUENCE OF 1-166 FROM N.A.

RX MEDLINE-91219421; PubMed-2023909;

RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,

RA Martal J.A.;

RT "The evolutionarily conserved Kruppel-associated box domain defines a

RT subfamily of eukaryotic multifingered proteins.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).

CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.

CC -1- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M27879; -; NOT ANNOTATED_CDS.

DR EMBL: M61867; AA58666.1; -

DR PIR: B39384; B39384.

DR HSP: P03001; 1TF6.

DR Genew: HGNC:12897; ZNF117.

DR MIM: 194624; -

DR InterPro: IPR001909; KRAB.

DR InterPro: IPR000822; Znf_C2H2.

DR Pfam: PF00096; zf-C2H2; 10.

DR PRODOM: PD000003; Znf_C2H2; 5.

DR SMART: SM00349; KRAB; 1.

DR SMART: SM00355; Znf_C2H2; 10.

DR PROSITE: PS50805; KRAB; 1.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.

DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.

DR Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;

KW Nuclear protein; Repeat.

FT DOMAIN 1 46 KRAB.

FT ZN_FING 116 417 ZINC_FINGERS.

FT ZN_FING 116 417 C2H2-TYPE (DEGENERATE).

FT ZN_FING 144 166 C2H2-TYPE.

FT ZN_FING 172 194 C2H2-TYPE (DEGENERATE).

FT ZN_FING 200 222 C2H2-TYPE.

FT ZN_FING 228 250 C2H2-TYPE.

FT ZN_FING 256 278 C2H2-TYPE.

FT ZN_FING 284 306 C2H2-TYPE.

FT ZN_FING 312 334 C2H2-TYPE.

FT ZN_FING 340 362 C2H2-TYPE.

FT ZN_FING 368 390 C2H2-TYPE (DEGENERATE).

FT ZN_FING 396 417 C2H2-TYPE.

FT CONFLICT 75 75 R -> G (IN REF. 2).

FT CONFLICT 78 79 G -> RH (IN REF. 2).

SO SEQUENCE 418 AA; 48941 MW; CDB554CB69B0B69 CRC64;

Query Match 1.9%; Score 8; DB 1; Length 418;

Best Local Similarity 100.0%; Pred. No. 3.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126

DB 21 PDLITCLE 28

RESULT 5

ID 2257_HUMAN STANDARD; PRT; 535 AA.

AC Q9Y201;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).

GN ZNF257 OR BMZF4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE-Bone marrow;

RX MEDLINE-20054457; PubMed-10585455;

RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,

RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;

RT "Molecular cloning of six novel Kruppel-like zinc finger genes from

```
RT hemopoietic cells and identification of a novel transregulatory
RT domain KRNB.*;
RL J. Biol. Chem. 274:35741-35748(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF070651; AAD20957.1; -.
DR HSSP; P08047; 1SF2.
DR Genew; HGNC:13498; ZNF257.
DR MIM; 606957; -.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf_C2H2; 10.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; ZnF_C2H2; 9.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZnF_C2H2; 10.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS50028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 4 75 KRAB.
FT ZN_FINGERS.
FT ZN_FING 173 509
FT ZN_FING 201 195 C2H2-TYPE.
FT ZN_FING 229 223 C2H2-TYPE.
FT ZN_FING 257 251 C2H2-TYPE.
FT ZN_FING 282 282 C2H2-TYPE (DEGENERATE).
FT ZN_FING 288 310 C2H2-TYPE.
FT ZN_FING 316 338 C2H2-TYPE.
FT ZN_FING 344 366 C2H2-TYPE.
FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).
FT ZN_FING 403 425 C2H2-TYPE.
FT ZN_FING 431 453 C2H2-TYPE.
FT ZN_FING 459 481 C2H2-TYPE.
FT ZN_FING 487 509 C2H2-TYPE.
FT ZN_FING 535 535 C2H2-TYPE.
SQ SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;
Query Match 1.9%; Score 8; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY .119 PDLITCLE 126
DB 50 PDLITCLE 57
RESULT 6
YE73_HUMAN STANDARD: PRT; 574 AA.
AC Q9P235;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA1473 (Fragment).
CN KIAA1473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
```

```
RP SEQUENCE FROM N.A.
RC TISUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB040906; BAA95997.1; -.
DR HSSP; P07248; 1ARD.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; zf_C2H2; 13.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; ZnF_C2H2; 2.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZnF_C2H2; 13.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS50028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT NON_TER 1 1
FT DOMAIN 15 86 KRAB.
FT ZN_FINGERS.
FT ZN_FING 184 542
FT ZN_FING 212 206 C2H2-TYPE.
FT ZN_FING 240 234 C2H2-TYPE.
FT ZN_FING 262 262 C2H2-TYPE.
FT ZN_FING 268 290 C2H2-TYPE.
FT ZN_FING 296 318 C2H2-TYPE.
FT ZN_FING 324 346 C2H2-TYPE.
FT ZN_FING 352 374 C2H2-TYPE.
FT ZN_FING 380 402 C2H2-TYPE.
FT ZN_FING 408 430 C2H2-TYPE.
FT ZN_FING 436 458 C2H2-TYPE.
FT ZN_FING 464 486 C2H2-TYPE.
FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).
FT ZN_FING 520 542 C2H2-TYPE.
SQ SEQUENCE 574 AA; 65951 MW; 482E5F1176BCA5B0 CRC64;
Query Match 1.9%; Score 8; DB 1; Length 574;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY .119 PDLITCLE 126
DB 61 PDLITCLE 68
RESULT 7
ZN85_HUMAN STANDARD: PRT; 595 AA.
AC Q03923;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 85 (zinc finger protein HPR4) (HNF1).
CN ZNF85.
```


OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 ON NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC TISSUE-PLACENTA;
 RA MEDLINE=99053537; PubMed=9839802;
 RA Poncelet D.A., Bellefroid E.J., Bastiaens P.V., Demotie M.A.,
 RA Marne J.C., Penderville H., Alam Y., Devos N., Lecocq P.J., Ogawa T.,
 RA Muller M., Marital J.A.;
 RT "functional analysis of ZNF85 KRAB zinc finger protein, a member of
 RT the highly homologous ZNF91 family.";
 RL DNA Cell Biol. 17:931-943(1998).
 RN [2]
 RP SEQUENCE OF 1-196 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 RA Marital J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN TESTICULAR
 CC TISSUES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U35376; AAA79179.1; -
 DR EMBL: M61866; AAA52689.1; -
 DR EMBL: M61868; AAA58671.1; -
 DR PIR: A39384; A39384.
 DR PIR: C39384; C39384.
 DR HSSP: P08048; ZNF.
 DR TRANSFAC: T04990; -
 DR Gene: HGNC:13160; ZNF85.
 DR MIM: 603899; -
 DR InterPro: IPR001099; KRAB.
 DR InterPro: IPR000822; ZNF_C2H2.
 DR Pfam: PF00096; zf-C2H2; 15.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR PRODOM: PD000003; Znf.C2H2; 13.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; ZNF_C2H2; 15.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 15.
 DR Transcription regulation: Zinc-finger; DNA-binding; Metal-binding;
 KM Nuclear protein; Repeat; Repressor.
 FT DOMAIN 4 75 KRAB.
 FT ZINC_FINGERS.
 FT ZN_FING 146 588 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 174 166 C2H2-TYPE.
 FT ZN_FING 202 224 C2H2-TYPE.
 FT ZN_FING 230 252 C2H2-TYPE.
 FT ZN_FING 258 280 C2H2-TYPE.
 FT ZN_FING 286 308 C2H2-TYPE.
 FT ZN_FING 314 336 C2H2-TYPE.
 FT ZN_FING 342 364 C2H2-TYPE.
 FT ZN_FING 370 392 C2H2-TYPE.
 FT ZN_FING 398 420 C2H2-TYPE.

FT ZN_FING 426 448 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 454 476 C2H2-TYPE.
 FT ZN_FING 482 504 C2H2-TYPE.
 FT ZN_FING 510 532 C2H2-TYPE.
 FT ZN_FING 538 560 C2H2-TYPE.
 FT ZN_FING 566 588 C2H2-TYPE.
 FT CONFLICT 84 84 R -> Q (IN REF. 2; AAA52689).
 FT CONFLICT 115 115 R -> I (IN REF. 2; AAA58671).
 FT CONFLICT 177 177 T -> R (IN REF. 2; AAA58671).
 FT CONFLICT 184 184 G -> R (IN REF. 2; AAA58671).
 SQ SEQUENCE 595 AA; 68718 MW; 44AA0A236D62D43B CRC64;
 QY 119 PDLITCLE 126 1.9%; Score 8; DB 1; Length 595;
 DB 50 PDLITCLE 57 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 8
 YEST_YEAST STANDARD; PRT; 624 AA.
 ID YEST_YEAST
 AC P40090;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 72.1 kDa protein in FTR1-SP715 intergenic region.
 GN YEX147C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Ayiles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Moseedale D., Nakahara K., Namath A., Norgren R., Oetner P., Oh C.,
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U18917; AAB6474.1; -
 DR SGD: S0000949; YER147C.
 DR InterPro: IPR000873; AMP-bind.
 DR PROSITE: PS00455; AMP_BINDING; 1.
 KM Hypothetical protein; Ligase.
 SQ SEQUENCE 624 AA; 72140 MW; 1269D931C656608 CRC64;
 QY 368 RIALKVEL 375 1.9%; Score 8; DB 1; Length 624;
 DB 285 RIALKVEL 292 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 9

DR PROSITE: PS00307; LECTIN_LEGUME_BETA; FALSE_NEG.
 DR PROSITE: PS00308; LECTIN_LEGUME_ALPHA; FALSE_NEG.
 KW LECTIN; signal; Multigene family; Manganese; Calcium; Glycoprotein.
 FT SIGNAL 1 21 OR 23 (POTENTIAL).
 FT CHAIN 22 265 NONSEED LECTIN.
 FT CARBOHYD 59 59 N-LINKED (GLCNAC.. .) (POTENTIAL).
 FT CARBOHYD 127 127 N-LINKED (GLCNAC.. .) (POTENTIAL).
 SQ SEQUENCE 265 AA; 28530 MW; BC3CC9BD7DCAB82 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 265;
 Best local similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 197 ILANGVL 203
 Db 49 ILANGVL 55
 RESULT 11
 XLR1_FUGRU STANDARD; PRT; 280 AA.
 AC Q9W6R5;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE X-linked juvenile retinoschisis protein precursor.
 GN XLR1.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 NCBI_TaxID=31033;
 RX MEDLINE-99263230; PubMed-10330123;
 RA Brunner B., Todd T., Lenzner S., Stout K., Schulz U., Ropers H.-H.,
 RA Kalschauer V.M.;
 RT "Genomic structure and comparative analysis of nine Fugu genes:
 RT conservation of synteny with human chromosome Xp22.2-p22.1.";
 RL Genome Res. 9:437-448(1999).
 CC -1- FUNCTION: MAY BE ACTIVE IN CELL ADHESION PROCESSES DURING RETINAL
 CC DEVELOPMENT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF146687; AAD28797.1; -;
 DR HSSP: P12259; 1CZT.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR001092; HLH_basic.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR SMART: SM00231; FA58C_1.
 DR PROSITE: PS01285; FA58C_1; 1.
 DR PROSITE: PS01286; FA58C_2; FALSE_NEG.
 KW Cell adhesion; signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 265 X-LINKED JUVENILE RETINOSCHISIS PROTEIN.
 FT DOMAIN 119 275 F5/8 TYPE C.
 FT DISULFID 119 275 BY SIMILARITY.
 SQ SEQUENCE 280 AA; 32032 MW; A0DF9A3222ED0167 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 280;
 Best local similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 351 PPIVARY 357
 Db 248 PPIVARY 254
 RESULT 12
 LEC_SOYBN STANDARD; PRT; 285 AA.
 ID LEC_SOYBN
 AC P05046;
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-JUL-1999 (Rel. 38, Last annotation update)
 DE Lectin precursor (Agglutinin) (SBA).
 GN LEL.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 NCBI_TaxID=3847;
 RX MEDLINE-84026469; PubMed-6313203;
 RA Vokkin I.O., Rhodes P.R., Goldberg R.B.;
 RT "A lectin gene insertion has the structural features of a
 RT transposable element.";
 RL Cell 34:1023-1031(1983).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RA MEDLINE-95226380; PubMed-7711015;
 RA Dessen A., Gupta D., Sabesan S., Brewer C.F., Sacchettini J.C.;
 RT "X-ray crystal structure of the soybean agglutinin cross-linked with
 RT a biantennary analog of the blood group I carbohydrate antigen.";
 RL Biochemistry 34:4933-4942(1995).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.52 ANGSTROMS).
 RA MEDLINE-98060759; PubMed-9398234;
 RA Olsen L.R., Dessen A., Gupta D., Sabesan S., Sacchettini J.C.,
 RA Brewer C.F.;
 RT "X-ray crystallographic studies of unique cross-linked lattices
 RT between four isomeric biantennary oligosaccharides and soybean
 RT agglutinin.";
 RL Biochemistry 36:15073-15080(1997).
 CC -1- FUNCTION: BINDS GALNAC AND GALACTOSE.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SIMILARITY: BELONGS TO THE LEGUMINOS LECTIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: K00821; AAA3983.1; -;
 DR PIR: S27365; S27365.
 DR PDB: 2SBA; 09-DEC-98.
 DR PDB: 1SBD; 22-APR-98.
 DR PDB: 1SBE; 22-APR-98.
 DR PDB: 1SBE; 22-APR-98.
 DR PDB: 1SBE; 22-APR-98.
 DR GlycoSuiteDB: P05046;
 DR InterPro: IPR000985; LECTIN_LEG.
 DR InterPro: IPR001220; LECTIN_LEG.
 DR Pfam: PF00139; LECTIN_LEG; 1.
 DR ProDom: PD000671; LECTIN_LEG; 1.
 DR ProDom: PD000711; LECTIN_LEG; 1.
 DR PROSITE: PS00307; LECTIN_LEGUME_BETA; 1.
 DR PROSITE: PS00308; LECTIN_LEGUME_ALPHA; 1.
 KW LECTIN; Glycoprotein; signal; 3D-structure.
 FT SIGNAL 1 32
 FT CHAIN 33 285 LECTIN.
 FT CARBOHYD 107 107 N-LINKED (GLCNAC.. .).

SQ SEQUENCE 285 AA; 30928 MW; B3704533C9315C52 CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 285;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 NESGDY 249
 |||||
 Db 146 NESGDY 152

RESULT 13
 NPL_HAEIN STANDARD; PRT; 293 AA.
 ID NPL_HAEIN
 AC P44539;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable N-acetylneuraminic lyase subunit (EC 4.1.3.3) (N-acetylneuraminic acid aldolase) (N-acetylneuraminic pyruvate lyase) (Nalase).
 GN NANA OR H10142.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Hemophilus.
 NC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sulten G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Uettermann T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., FINE L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., RAE Grahm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae R4.";
 RT Science 269:496-512(1995).
 RL -1- CATALYTIC ACTIVITY: N-acetylneuraminic acid - N-acetyl-D-mannosamine + pyruvate.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DHDS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U37200; AAC21814.1; -
 DR HSSP: P06995; INAL.
 DR TIGR: H10142; -
 DR InterPro: IPR002220; DHDS.
 DR InterPro: IPR005264; Nana.
 DR Pfam: PF00701; DHDS.1.
 DR PRINTS: PR00146; DHPICNTHASE.
 DR PRODOM: PD001859; DHDS.1.
 DR TIGRFAMS: TIGR00683; nana.1.
 DR PROSITE: PS00665; DHDS.1; 1.
 DR PROSITE: PS00666; DHDS.2; 1.
 KW Lyase; Complete proteome.
 FT ACT_SITE 164 164 BY SIMILARITY.
 SQ SEQUENCE 293 AA; 32564 MW; ACAV550DA5D93D33 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 293;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 EGIANG 201
 |||||
 Db 243 EGIANG 249

RESULT 14
 OM40_CAEEL STANDARD; PRT; 301 AA.
 ID OM40_CAEEL
 AC Q18090;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable mitochondrial import receptor subunit TOM40 homolog (Translocase of outer membrane 40 kDa subunit homolog).
 GN C18B9.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Felodierinae; Caenorhabditis.
 NC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2.
 RA Sims M.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL FOR THE IMPORT OF PROTEIN PRECURSORS INTO THE MITOCHONDRIA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial outer membrane (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE TOM40 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z70034; CAA93850.1; -
 DR Wormpep: C18B9.6; CE05298.
 DR Hypothetical protein; Transport; Protein transport; Outer membrane; Mitochondrion; Transmembrane.
 KW Mitochondrion; Transmembrane.
 SQ SEQUENCE 301 AA; 32386 MW; 4247A3593B5EBC57 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 301;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 GGIISVL 186
 |||||
 Db 181 GGIISVL 187

RESULT 15
 Y073_LISMO STANDARD; PRT; 322 AA.
 ID Y073_LISMO
 AC P58588;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein lmo2473.
 GN LMO2473.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 NC NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Bergeul L., Buchrieser C., Rusnok C., Amend A., Bagueri F., Berche P., Bloeker H., Brandt P., Chakraborty T., Chardot A., Chetoui F., Couve E., de Darvar A., Denoux P.,

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:16:36 ; Search time 22.3469 Seconds
(without alignments)
1305.856 Million cell updates/sec

Title: US-10-003-132-4-COPY_26_244

Perfect score: 219
Sequence: 1 EELDGCGCHITVSQDSGTWT.....NGSLRHGSLSEKRELFETTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 306882

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

1: A.Geneseq.101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	67.6	503	21	AA19126
2	28	12.8	487	22	AAU00628
3	28	12.8	539	22	AAU00630
4	28	12.8	586	22	AAU00629
5	20	9.1	398	23	AAE22715
6	20	9.1	398	23	AAU79459
7	8	3.7	101	22	AAU19467
8	8	3.7	101	23	ABP51369
9	8	3.7	101	23	ABP51461
10	8	3.7	109	22	AA198946

11	8	3.7	117	22	AA192967
12	8	3.7	117	23	ABP51401
13	8	3.7	128	22	AA179595
14	8	3.7	193	22	AA140732
15	8	3.7	273	22	AB160044
16	8	3.7	353	20	AA139779
17	8	3.7	364	22	AA194710
18	8	3.7	493	22	AB162866
19	8	3.7	694	21	AA199661
20	7	3.2	115	22	AB127684
21	7	3.2	126	22	AA190976
22	7	3.2	204	23	AA183377
23	7	3.2	206	21	AA196283
24	7	3.2	217	23	AB126698
25	7	3.2	220	22	AB121553
26	7	3.2	263	15	AA163754
27	7	3.2	275	22	AA186528
28	7	3.2	297	22	AB112597
29	7	3.2	312	22	AA114605
30	7	3.2	350	21	AA121445
31	7	3.2	350	22	AA193217
32	7	3.2	350	22	AA193281
33	7	3.2	350	22	AA194684
34	7	3.2	359	20	AA181359
35	7	3.2	385	22	AA181097
36	7	3.2	414	22	AA110702
37	7	3.2	430	23	AB128277
38	7	3.2	464	22	AA193631
39	7	3.2	474	22	AB162780
40	7	3.2	474	22	AB162780
41	7	3.2	523	18	AA138187
42	7	3.2	524	21	AA128588
43	7	3.2	610	22	AA134821
44	7	3.2	639	22	AB117654
45	7	3.2	641	22	AB158034

ALIGNMENTS

RESULT 1	AA19126
ID	AA19126 standard; Protein; 503 AA.
XX	
AC	AA19126;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Polypeptide isolated, from lymph node stromal cells of fsn -/- mice.
XX	
KW	Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW	immune system disorder; cancer; viral infection;
KW	blood vessel growth; tumour necrosis factor disorder; arthritis;
KW	inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW	cardiac failure.
XX	
OS	Mus sp.
XX	
PN	W0200058463-A1.
XX	
PD	05-OCT-2000.
XX	
PF	18-FEB-2000; 2000MO-NZ00015.
XX	
PR	25-MAR-1999; 99US-0276268.
XX	
PR	26-AUG-1999; 99US-0383386.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	
PI	Strachan L, Sleeman M, Abernethy N, Orrust R, Kumble KD;
XX	Murison JG;
DR	WPI; 2000-664924/64.

DR N-PSDB: AAA96736.
XX Polypeptide expressed in mammalian fkn -/- lymph node stromal cells,
PT useful for modulating growth of blood cells, for treating inflammatory
PT and tumour necrosis factor-mediated disorders, cancer and viral
PT disorders -
XX
PS Claim 1; Page 68-69; 75pp; English.
XX
XX The present sequence represents a polypeptide sequence which is
CC isolated from lymph node stromal cells of fkn -/- mice. The
CC polynucleotides and their polypeptides are useful for treating an
CC inflammatory disorder, disorder of immune system and cancer selected
CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
CC viral disorder, in particular HIV infection and for modulating the
CC growth of blood vessels. The polypeptides are useful for treating a
CC tumour necrosis factor (TNF) mediated disorder, such as those selected
CC from arthritis, inflammatory bowel disease and cardiac failure and a
CC fibroblast growth factor-mediated disorder. It is also useful in assays
CC to determine biological activity, to raise antibodies, to isolate
CC corresponding ligands or receptors, to quantify levels of protein or
CC cognate corresponding ligand or receptors, as anti-inflammatory agents,
CC and in compositions for the treatment of skin, connective tissue and
CC immune system diseases. The polynucleotide is useful as marker for
CC tissue, as a chromosome marker or tags in the identification of a
CC genetic disorder.
XX
SO Sequence 503 AA;

Query Match 67.6%; Score 148; DB 21; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.4e-18;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EELGDCGHIYVSDSGTWSKNTKYPNTYVCEKIIIVPGKRLILRLGLNIESKTC 60
DB 26 EELGDCGHIYVSDSGTWSKNTKYPNTYVCEKIIIVPGKRLILRLGLNIESKTC 85
OY 61 ASDYLLFSSATQYCPYCGSNAVPEKLRNSNEVTVLFRSGSHIGKFLITYASSDHPD 120
DB 86 ASDYLLFSSATQYCPYCGSNAVPEKLRNSNEVTVLFRSGSHIGKFLITYASSDHPD 145
OY 121 LITCLERSHYFEKYSKFCPCAGCDIA 148
DB 146 LITCLERSHYFEKYSKFCPCAGCDIA 173

RESULT 2

AAU00628 ID AAU00628 standard; Protein: 487 AA.

AC AAU00628;

DT 29-AUG-2001 (first entry)

XX Novel human protein (NHP) sequence #1.

DE Novel human protein (NHP) sequence #1.
XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX

PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB: AAS00613.
XX

PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 4; Page 26; 33pp; English.
XX

XX The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
SO Sequence 487 AA;

Query Match 12.8%; Score 28; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.7e-19;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 SCSHISGRGFLITYASSDHPDLITCLER 127
DB 82 SCSHISGRGFLITYASSDHPDLITCLER 109

RESULT 3

AAU00630 ID AAU00630 standard; Protein: 539 AA.

AC AAU00630;

DT 29-AUG-2001 (first entry)

XX Novel human protein (NHP) sequence #3.

DE Novel human protein (NHP) sequence #3.
XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB: AAS00615.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders

PT or diseases -
XX
PS Claim 5; Page 29-30; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
CC
SQ Sequence 539 AA;

Query Match 12.8%; Score 28; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 4e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 GSGHISGRGFLTYASSDHPDLITCLER 127
Db 134 GSGHISGRGFLTYASSDHPDLITCLER 161
|||||

RESULT 4
AAU00629
ID AAU00629 standard; Protein; 586 AA.
XX
AC AAU00629;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #2.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000MO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00614.
XX
XX Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 2; Page 27-28; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing

CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
CC
SQ Sequence 586 AA;

Query Match 12.8%; Score 28; DB 22; Length 586;
Best Local Similarity 100.0%; Pred. No. 4.3e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 GSGHISGRGFLTYASSDHPDLITCLER 127
Db 181 GSGHISGRGFLTYASSDHPDLITCLER 208
|||||

RESULT 5
AAE22715
ID AAE22715 standard; Protein; 398 AA.
XX
AC AAE22715;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy1 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW ankyrotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnerary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cyostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX
PN WO200222815-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001MO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX WPI; 2002-393966/42.
DR N-PSDB; AAD35992.
XX
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects -
XX
PS Claim 3; Page 123-125; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopenia, plaitic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders. Infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye diseases. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy1
CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

Sequence 398 AA;

Query Match 9.1%; Score 20; DB 23; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 100 SCSHISGRGFLTYASDHP 119
|||||

Db 100 SCSHISGRGFLTYASDHP 119

RESULT 6

ID AAU79459 standard; Protein: 398 AA.

AC AAU79459;

DT 15-JUL-2002 (first entry)

DE Human Neuropilin-Hy1.

KW Human; neuropilin-Hy1; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.

OS Homo sapiens.

PN WO200222780-A2.

PD 21-MAR-2002.

PE 11-SEP-2001; 2001WO-US28590.

PR 11-SEP-2000; 2000US-0659671.

PA (TANG/) TANG T Y.

PI Tang TY;
XX WPI: 2002-351881/38.
DR N-PSDB; ABR49565.

PT New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation
PS Claim 3; Page 118-120; 144pp; English.

XX The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischaemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy1.

Sequence 398 AA;

Query Match 9.1%; Score 20; DB 23; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.6e-11;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 100 SCSHISGRGFLTYASDHP 119
|||||

Db 100 SCSHISGRGFLTYASDHP 119

RESULT 7

ID AAU19467 standard; Protein: 101 AA.

AC AAU19467;

DT 04-DEC-2001 (first entry)

DE Human diagnostic and therapeutic polypeptide (DITHP) #53.

KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder.

OS Homo sapiens.

PN WO200162927-A2.

XX 30-AUG-2001.
 XX 21-FEB-2001: 2001MO-US06059.
 XX
 XX
 PR 24-FEB-2000: 2000US-0184693.
 PR 24-FEB-2000: 2000US-0184697.
 PR 24-FEB-2000: 2000US-0184698.
 PR 24-FEB-2000: 2000US-0184768.
 PR 24-FEB-2000: 2000US-0184769.
 PR 24-FEB-2000: 2000US-0184770.
 PR 24-FEB-2000: 2000US-0184771.
 PR 24-FEB-2000: 2000US-0184772.
 PR 24-FEB-2000: 2000US-0184773.
 PR 24-FEB-2000: 2000US-0184774.
 PR 24-FEB-2000: 2000US-0184776.
 PR 24-FEB-2000: 2000US-0184777.
 PR 24-FEB-2000: 2000US-0184797.
 PR 24-FEB-2000: 2000US-0184813.
 PR 24-FEB-2000: 2000US-0184837.
 PR 24-FEB-2000: 2000US-0184841.
 PR 24-FEB-2000: 2000US-0185213.
 PR 24-FEB-2000: 2000US-0185215.
 PR 12-MAY-2000: 2000US-0203785.
 PR 15-MAY-2000: 2000US-0204226.
 PR 16-MAY-2000: 2000US-0204525.
 PR 16-MAY-2000: 2000US-0204821.
 PR 16-MAY-2000: 2000US-0204908.
 PR 16-MAY-2000: 2000US-0205232.
 PR 17-MAY-2000: 2000US-0204815.
 PR 17-MAY-2000: 2000US-0204863.
 PR 17-MAY-2000: 2000US-0205221.
 PR 17-MAY-2000: 2000US-0205285.
 PR 17-MAY-2000: 2000US-0205286.
 PR 17-MAY-2000: 2000US-0205287.
 PR 17-MAY-2000: 2000US-0205323.
 PR 17-MAY-2000: 2000US-0205324.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'Sa SA, Amshay S, Dahl CR, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Ross BH, Russo FD, Stockreiter TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
 XX
 DR WPI: 2001-502867/55.
 DR N-PSDB: AAS31038.
 XX
 XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics.
 PT
 PT
 XX
 XX Claim 27: Page 430: 522pp: English.
 XX
 CC The invention relates to polynucleotides (I) encoding diagnostic and
 CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
 CC and proteins involved in growth and development and receptors. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate DITHP expression. For example, (I) and
 CC (II) may be used to treat disorders associated with decreased polypeptide
 CC expression by rectifying mutations or deletions in a patient's genome,
 CC that affect the activity of the DITHPs, by expressing inactive proteins
 CC or supplementing the patient's own production of them. (I) and (II)
 CC may be used to treat diseases, for example, cell proliferative disorder,
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
 CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
 CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC its complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative

CC therapy. (II) may also be used as antigens in the production of
 CC antibodies against DITHPs and in assays to identify modulators of DITHP
 CC expression and activity. The anti-DITHP antibodies and antagonists may
 CC also be used to down regulate expression and activity. The anti-DITHP
 CC antibodies may also be used as diagnostic agents for detecting the
 CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent
 CC assay (ELISA)). AA019415-AA019625 represent human diagnostic and
 CC therapeutic (DITHP) polypeptides of the invention.
 XX
 XX Sequence 101 AA:
 Query Match 3.7%; Score 8; DB 22; Length 101;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 119 PDLITCLE 126
 Db 59 PDLITCLE 66
 RESULT 8
 ABP51369
 ID ABP51369 standard; Protein; 101 AA.
 XX
 AC ABP51369;
 XX
 DT 03-SEP-2002 (first entry)
 XX
 DE Human MDDT SEQ ID NO 391.
 XX
 XX Human; MDDT; disease detection and treatment molecule polynucleotide;
 KW Proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;
 KW hepatocytic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;
 KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antigout;
 KW neuroprotective; antineumatic; antiarthritic.
 XX
 OS Homo sapiens.
 XX
 PN WO200240715-A2.
 XX
 PD 23-MAY-2002.
 XX
 PF 06-SEP-2001: 2001MO-US27628.
 XX
 XX 06-SEP-2000: 2000US-230505P.
 PR 06-SEP-2000: 2000US-230514P.
 PR 06-SEP-2000: 2000US-230515P.
 PR 06-SEP-2000: 2000US-230517P.
 PR 06-SEP-2000: 2000US-230518P.
 PR 06-SEP-2000: 2000US-230519P.
 PR 06-SEP-2000: 2000US-230595P.
 PR 06-SEP-2000: 2000US-230597P.
 PR 06-SEP-2000: 2000US-230598P.
 PR 06-SEP-2000: 2000US-230599P.
 PR 06-SEP-2000: 2000US-230610P.
 PR 06-SEP-2000: 2000US-230865P.
 PR 06-SEP-2000: 2000US-230988P.
 PR 06-SEP-2000: 2000US-230989P.
 PR 07-SEP-2000: 2000US-230951P.
 PR 07-SEP-2000: 2000US-231163P.
 PR 07-SEP-2000: 2000US-231167P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
 PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
 XX
 DR WPI: 2002-527544/56.

DR N-PSDB; ABQ72586.

XX Novel human disease detection and treatment polypeptide, useful in

PT diagnosis, prevention or treatment of cell proliferative disorders e.g.

PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder

PT e.g. AIDS

PS Claim 14; Page 541-542; 618pp; English.

XX The invention relates to an isolated human disease detection and

XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a

CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the

CC specification, a naturally occurring polypeptide comprising a sequence

CC having at least 90% identity to (I) or a biologically active or

CC immunogenic fragment of (I). (I) is useful for screening a compound for

CC effectiveness as an agonist or antagonist, for screening a compound that

CC specifically binds (I) or modulates the activity of (I), and for

CC preparing a polyclonal or monoclonal antibody by hybridoma technology.

CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for

CC screening a compound for effectiveness in altering expression of a target

CC polynucleotide comprising. Oligonucleotides and antibodies are useful for

CC detecting MDP in a sample or for assessing toxicity of a test compound,

CC in a diagnostic test for a condition or a disease associated with the

CC expression of MDP in a biological sample, for detecting (I) in a sample,

CC and for purifying (I) from a sample. A composition comprising (I), an

CC agonist or antagonist is useful for treating a disease or condition.

CC associated with decreased or increased expression of functional MDP.

CC (I) or (II) are useful for diagnosing, treating or preventing disorders

CC associated with aberrant expression of MDP, where the disorders are

CC selected from a cell proliferative disorder such as arteriosclerosis,

CC cirrhosis, hepatitis, psoriasis, and cancer and an

CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,

CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or

CC rheumatoid arthritis. (II) are useful for creating knock-in humanised

CC animals or transgenic animals to model human diseases. In somatic or

CC germ-line gene therapy, to generate a transcript image of a tissue or cell

CC type, for detecting differences in the chromosomal location due to

CC translocation or inversion among normal, carrier or affected individuals

CC and as hybridisation probes for mapping naturally occurring genomic

CC sequences.

XX

SQ Sequence 101 AA;

Query Match 3.7%; Score 8; DB 23; Length 101;

Best Local Similarity 100.0%; Pred. No. 6.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLINCLE 126

DB 59 PDLINCLE 66

RESULT 9

ABP51461

ID ABP51461 standard; Protein; 101 AA.

XX

AC ABP51461;

XX

DT 03-SEP-2002 (first entry)

XX

DE Human MDP ID NO 483.

XX

Human; MDP; disease detection and treatment molecule polynucleotide;

XX proliferative disorder; hepatitis; psoriasis; cancer; AIDS;

XX autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;

XX rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;

XX hepatotropic; antiinflammatory; antipsoriatic; cycostatic; anti-HIV;

XX antiallergic; antianaemic; antiaschmatic; antiatherosclerotic; anti-gout;

XX neuroprotective; antineumatic; antiarthritic.

OS Homo sapiens.

XX

PN WO200240715-A2.

XX

PD 23-MAY-2002.

XX

PF 06-SEP-2001; 2001WO-US27628.

XX

PR 06-SEP-2000; 2000US-230505P.

PR 06-SEP-2000; 2000US-230514P.

PR 06-SEP-2000; 2000US-230515P.

PR 06-SEP-2000; 2000US-230517P.

PR 06-SEP-2000; 2000US-230518P.

PR 06-SEP-2000; 2000US-230519P.

PR 06-SEP-2000; 2000US-230595P.

PR 06-SEP-2000; 2000US-230597P.

PR 06-SEP-2000; 2000US-230598P.

PR 06-SEP-2000; 2000US-230599P.

PR 06-SEP-2000; 2000US-230610P.

PR 06-SEP-2000; 2000US-230655P.

PR 06-SEP-2000; 2000US-230888P.

PR 06-SEP-2000; 2000US-230898P.

PR 07-SEP-2000; 2000US-230951P.

PR 07-SEP-2000; 2000US-231163P.

PR 07-SEP-2000; 2000US-231167P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JT;

PI Jones AL, Yu JY, Wright RJ, Gletzen D, Liu TF, Yap PE, Dahl CR;

PI Momiyama NG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;

PI Gerlitz EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;

PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;

XX

DR WPI; 2002-527544/56.

XX

DR N-PSDB; ABQ72678.

XX

PT Novel human disease detection and treatment polypeptide, useful in

PT diagnosis, prevention or treatment of cell proliferative disorders e.g.

PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder

PT e.g. AIDS

PS Claim 14; Page 603; 618pp; English.

XX

The invention relates to an isolated human disease detection and

CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a

CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the

CC specification, a naturally occurring polypeptide comprising a sequence

CC having at least 90% identity to (I) or a biologically active or

CC immunogenic fragment of (I). (I) is useful for screening a compound for

CC effectiveness as an agonist or antagonist, for screening a compound that

CC specifically binds (I) or modulates the activity of (I), and for

CC preparing a polyclonal or monoclonal antibody by hybridoma technology.

CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for

CC screening a compound for effectiveness in altering expression of a target

CC polynucleotide comprising. Oligonucleotides and antibodies are useful for

CC detecting MDP in a sample or for assessing toxicity of a test compound,

CC in a diagnostic test for a condition or a disease associated with the

CC expression of MDP in a biological sample, for detecting (I) in a sample,

CC and for purifying (I) from a sample. A composition comprising (I), an

CC agonist or antagonist is useful for treating a disease or condition.

CC associated with decreased or increased expression of functional MDP.

CC (I) or (II) are useful for diagnosing, treating or preventing disorders

CC associated with aberrant expression of MDP, where the disorders are

CC selected from a cell proliferative disorder such as arteriosclerosis,

CC cirrhosis, hepatitis, psoriasis, and cancer and an

CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,

CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or

CC rheumatoid arthritis. (II) are useful for creating knock-in humanised

CC animals or transgenic animals to model human diseases. In somatic or

CC germ-line gene therapy, to generate a transcript image of a tissue or cell

CC type, for detecting differences in the chromosomal location due to

CC translocation or inversion among normal, carrier or affected individuals

CC and as hybridisation probes for mapping naturally occurring genomic

CC sequences.

XX

SQ Sequence 101 AA;
 Query Match 3.7%; Score 8; DB 23; Length 101;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 119 PDLITCLE 126
 |||||||
 DB 59 PDLITCLE 66
 RESULT 10
 AAM38946
 ID AAM38946 standard; Protein; 109 AA.
 AC AAM38946;
 XX
 XX 22-OCT-2001 (first entry)
 DT
 XX Human polypeptide SEQ ID NO 2091.
 DE
 XX Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CMS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 OS Homo sapiens.
 XX
 XX WO200153312-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX 26-DEC-2000; 2000MO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA
 XX (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 PI
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI58102.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Example 3; SEQ ID NO 2091; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA442213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.

CC Note: the sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 109 AA;
 Query Match 3.7%; Score 8; DB 22; Length 109;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 119 PDLITCLE 126
 |||||||
 DB 82 PDLITCLE 89
 RESULT 11
 AAB92967
 ID AAB92967 standard; Protein; 117 AA.
 AC AAB92967;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 XX Human protein sequence SEQ ID NO:11667.
 DE
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW
 KW Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA
 XX (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 PI
 XX WPI; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID-11667; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SO Sequence 117 AA;

Query Match 3.7%; Score 8; DB 22; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLTICLE 126
|||||
Db 81 PDLTICLE 88

RESULT 12
AAB51401
ID AAB51401 standard; Protein: 117 AA.

XX AAB51401;
AC AAB51401;
DT 03-SEP-2002 (first entry)
DE Human MDDF SEQ ID NO 423.

XX Human; MDDF; disease detection and treatment molecule polynucleotide;
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW rheumatoid arthritis; transgenic; gene therapy; arteriosclerosis;
KW hepatotropic; antiinflammatory; antipruritic; cycostatic; anti-HIV;
KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; antigout;
KW neuroprotective; antirheumatic; antirheumatic;
XX Homo sapiens.
OS
XX MO200240715-A2.
PN
XX 23-MAY-2002.
PD
XX 06-SEP-2001; 2001WO-US27628.
PE
XX 06-SEP-2000; 2000US-230505P.
PR 06-SEP-2000; 2000US-230514P.
PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230595P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230598P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230655P.
PR 06-SEP-2000; 2000US-230988P.
PR 06-SEP-2000; 2000US-230989P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JT;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
XX
XX WPI: 2002-527544/56.
DR N-PSDB; ABO72618.
DR
XX
XX Novel human disease detection and treatment polypeptide, useful in
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder

PT e.g. AIDS -
XX
PS Claim 14; Page 562; 618pp; English.

CC The invention relates to an isolated human disease detection and
CC treatment (MDDF) polypeptide (I) selected from a polypeptide having a
CC sequence selected from 254 sequences (AAB51231-AAB51484) given in the
CC specification, a naturally occurring polypeptide comprising a sequence
CC having at least 90% identity to (I) or a biologically active or
CC immunogenic fragment of (I). (I) is useful for screening a compound for
CC effectiveness as an agonist or antagonist, for screening a compound that
CC specifically binds (I) or modulates the activity of (I), and for
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
CC screening a compound for effectiveness in altering expression of a target
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDDF in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDDF in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDDF.
CC (I) or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDDF, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases. In somatic or
CC germ-line gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.

SO Sequence 117 AA;

Query Match 3.7%; Score 8; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLTICLE 126
|||||
Db 90 PDLTICLE 97

RESULT 13
AAM79595
ID AAM79595 standard; Protein: 128 AA.

XX AAM79595;
AC AAM79595;
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human protein SEQ ID NO 3241.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.
OS
XX MO200157190-A2.
PN
XX 09-AUG-2001.
PD
XX 05-FEB-2001; 2001WO-US04098.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0726422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhang QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK52728.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 293; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibit activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAK80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 128 AA;

Query Match 3.7%; Score 8; DB 22; Length 128;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 101 PDLITCLE 108

RESULT 14
AAM40732
ID AAM40732 standard; Protein; 193 AA.
XX
XX AAM40732;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 5663.
DE
XX
XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR
XX
XX 25-APR-2000; 2000US-0552317.
PR

PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI59888.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5663; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAK42213) with neotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 193 AA;

Query Match 3.7%; Score 8; DB 22; Length 193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 100 PDLITCLE 107

RESULT 15
ABG60044
ID ABG60044 standard; Protein; 273 AA.
XX
XX ABG60044;
AC
XX
XX 30-JUL-2002 (first entry)
DT
XX
XX Human DTHP polypeptide #102.
DE
XX
XX Human; DTHP; diagnostic and therapeutic polypeptide; bone; testis; skin;
KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;
KW inflammatory disorder; viral infection; bacterial infection; seizure;
KW fungal infection; parasitic infection; developmental disorder; breast;
KW endocrine disorder; metabolic disorder; neurological disorder; cervix;
KW gastrointestinal disorder; transport disorder; gene therapy; kidney;
KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
KW thymus.
XX
XX Homo sapiens.
OS
XX
XX WO200220754-A2.
PN
XX
XX 14-MAR-2002.
PD

```

XX 29-AUG-2001; 2001WO-US27127.
PE
XX
PR 05-SEP-2000; 2000US-229747P.
PR 05-SEP-2000; 2000US-229748P.
PR 05-SEP-2000; 2000US-229749P.
PR 05-SEP-2000; 2000US-229750P.
PR 05-SEP-2000; 2000US-229751P.
PR 05-SEP-2000; 2000US-230583P.
PR 05-SEP-2000; 2000US-230585P.
PR 05-SEP-2000; 2000US-230515P.
PR 05-SEP-2000; 2000US-230517P.
PR 05-SEP-2000; 2000US-230518P.
PR 05-SEP-2000; 2000US-230519P.
PR 05-SEP-2000; 2000US-230595P.
PR 05-SEP-2000; 2000US-230597P.
PR 05-SEP-2000; 2000US-230598P.
PR 05-SEP-2000; 2000US-230599P.
PR 05-SEP-2000; 2000US-230610P.
PR 05-SEP-2000; 2000US-230865P.
PR 05-SEP-2000; 2000US-230988P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Datto A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
XX MPI: 2002-383054/41.
XX N-PSDB; ABK71635.
XX
XX An isolated polynucleotide useful in diagnostics and therapeutics -
XX
XX Claim 29; Page 588-589; 686pp; English.
XX
XX The invention relates to human diagnostic and therapeutic (dthp)
XX polynucleotides and their associated polypeptides (DTHP polypeptides).
XX The sequences of the invention are used in the treatment and diagnosis of
XX cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers
XX (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,
XX cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or
XX thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,
XX psoriasis, osteoporosis), viral infections, bacterial infections, fungal
XX infections, parasitic infections, developmental disorders (e.g. anaemia,
XX epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),
XX endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders
XX (e.g. obesity, diabetes), neurological disorders (e.g. stroke,
XX amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal
XX disorders (e.g. ulcerative colitis, lysinuria) and transport disorders
XX (e.g. myotonic dystrophy, catatonla, peripheral neuropathy). Sequences
XX ABG39943-ABG60220 represent human DTHP polypeptides of the invention.
XX
XX
SQ Sequence 273 AA:

```

```

Query Match 3.7%; Score 8; DB 23; Length 273;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 119 PDLITCLE 126
   |||||
DB 101 PDLITCLE 108

```

```

Search completed: May 15, 2003, 13:22:40
Job time : 23.3469 secs

```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:21:41 ; Search time 19.2527 Seconds

(without alignments)
1093.529 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EEKGDGCGHIVTSQDSGTM.....NGVLSRHGSLSEKRFLETPP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 223049

Minimum DB seq length: 100
Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : PIR.73:*

- 1: PIR1:*
- 2: PIR2:*
- 3: PIR3:*
- 4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.7	154	2 D39384	finger protein HTF
2	8	3.7	163	2 B39384	finger protein HTF
3	8	3.7	195	2 E39384	finger protein HTF
4	8	3.7	196	2 A39384	finger protein HTF
5	8	3.7	440	1 R6ECUC	regulatory protein
6	8	3.7	440	2 D91204	regulator of uHPT
7	8	3.7	440	2 F86050	transducin homolog
8	8	3.7	519	2 A49367	transcription repr
9	8	3.7	595	2 G02075	probable ATP /GTP
10	8	3.7	135	2 D81337	hypothenase (EC 1.
11	8	3.7	152	2 JE0031	riboflavin synthas
12	8	3.7	203	2 AG2720	beta-lactamase (EC
13	8	3.7	203	2 AG2720	conserved hypochet
14	8	3.7	204	2 AG2720	phosphoribosylglyc
15	8	3.7	291	2 JP0074	recf protein - Myc
16	8	3.7	302	2 T50737	keratin, type I cy
17	8	3.7	314	2 C90256	cysteine proteinas
18	8	3.7	320	2 T27372	hypothetical prote
19	8	3.7	384	1 E69685	phosphoribosylglyc
20	8	3.7	385	2 S70984	keratin, type I cy
21	8	3.7	419	2 A25438	hypothetical prote
22	8	3.7	494	2 S51117	scarcrow-like pro
23	8	3.7	507	2 T00682	hypothetical prote
24	8	3.7	573	2 T51239	hypothetical prote
25	8	3.7	578	2 B71431	probable phosphos
26	8	3.7	578	2 S51379	DNA helicase reco
27	8	3.7	610	1 BVECR0	ATP-dependent DNA
28	8	3.7	611	2 H91222	ATP-dependent DNA
29	8	3.7	611	2 F86069	ATP-dependent DNA

30	7	3.2	642	2 S59306	probable membrane
31	6	2.7	102	2 A72739	probable DNA-dirc
32	6	2.7	110	2 S17709	kgf protein - Etw
33	6	2.7	111	2 T40753	very hypothetical
34	6	2.7	117	2 T35913	hypothetical prote
35	6	2.7	122	2 A84989	50S ribosomal prot
36	6	2.7	122	2 AC2172	hypothetical prote
37	6	2.7	123	2 AC2259	hypothetical prote
38	6	2.7	126	2 S30707	hypothetical 14.5k
39	6	2.7	126	2 AE0918	probable membrane
40	6	2.7	127	2 B85638	hypothetical prote
41	6	2.7	129	2 G90761	hypothetical prote
42	6	2.7	129	2 B85625	hypothetical prote
43	6	2.7	132	2 A82847	hypothetical prote
44	6	2.7	134	2 E75356	hypothetical prote
45	6	2.7	137	1 PWYCE	H+-transporting tw

ALIGNMENTS

RESULT 1
D39384
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 08-Dec-2000
C:Accession: D39384
R:Belletroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Kriepel-associated box domain defines a subfam
A:Accession: D39384; MUID:91219421; PMID:2023909
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <BRL>
A:Cross-references: GB:M61869; NID:g184449; PID:g184450
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 3.7%; Score 8; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 8 PDLITCLE 15

RESULT 2
B39384
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 08-Dec-2000
C:Accession: B39384
R:Belletroid, E.J.; Poncelet, D.A.; Lecocq, P.J.; Revelant, O.; Martial, J.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 3608-3612, 1991
A:Title: The evolutionarily conserved Kriepel-associated box domain defines a subfam
A:Reference number: A39384; MUID:91219421; PMID:2023909
A:Accession: B39384
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-163 <AAA>
A:Cross-references: GB:M61867
C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C:Keywords: DNA binding; zinc finger

Query Match 3.7%; Score 8; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
DB 18 PDLITCLE 25

A:Accession: F86050
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-440 <STO>
 A:Cross-references: GB:AE005174; NID:912518499; PIDN:AAG5866.1; GSPDB:GN00145; UWGP:Z51
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: uhpC
 C:Superfamily: hexose phosphate transport protein uhpT

Query Match
 Best Local Similarity 3.7%; Score 8; DB 2; Length 440;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 LANGVLSR 205
 Db 58 LANGVLSR 65

RESULT 8
 A49367
 transducin homolog sazd - human
 C:Species: Homo sapiens (man)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-May-2000
 C:Accession: A49367
 R:Weinstein-Saslow, D.L.; Germino, G.G.; Somlo, S.; Reeders, S.T.
 A:Title: A transducin-like gene maps to the autosomal dominant polycystic kidney disease
 A:Reference number: A49367; MUID:94140377; PMID:8307562
 A:Accession: A49367
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-519 <WEI>
 A:Cross-references: GB:U02609; NID:9414535; PIDN:AAA18945.1; PID:9414536
 C:Genetics:
 A:Gene: sazd
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:77-110/Domain: WD repeat homology <WD1>
 F:258-291/Domain: WD repeat homology <WD2>
 F:361-394/Domain: WD repeat homology <WD3>

Query Match
 Best Local Similarity 3.7%; Score 8; DB 2; Length 519;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 LFFSSATD 72
 Db 53 LFFSSATD 60

RESULT 9
 G02075
 transcription repressor zinc finger protein 85 - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
 C:Accession: G02075
 R:Poncellet, D.A.
 A:Submitted to the EMBL Data Library, September 1995
 A:Reference number: G09169
 A:Accession: G02075
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-595 <PON>
 A:Cross-references: EMBL:U05376; NID:91017721; PIDN:AAA9179.1; PID:91017722
 C:Genetics:
 A:Gene: GDB:ZNF85
 A:Cross-references: GDB:133279
 A:Map position: 19p12-19p12
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match
 Best Local Similarity 3.7%; Score 8; DB 2; Length 595;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
 Db 50 PDLITCLE 57

RESULT 10
 D81337
 probable ATP/GTP-binding protein Cj0668 [imported] - Campylobacter jejuni (strain NC
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: D81337
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanyilet, A.; Whitehead, S.; Bar
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: D81337
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-135 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB72945.1; PID:9696
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0668

Query Match
 Best Local Similarity 3.2%; Score 7; DB 2; Length 135;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 EGLLANG 201
 Db 80 EGLLANG 86

RESULT 11
 JE0031
 nitrogenase (EC 1.18.6.1) molybdenum-iron protein nifX - Rhodobacter capsulatus
 C:Species: Rhodobacter capsulatus
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 19-Jan-2001
 C:Accession: JE0031; S04417
 R:Morono-Vivian, C.; Schmehl, M.; Masepohl, B.; Arnold, W.; Klipp, W.
 Mol. Gen. Genet. 216, 353-363, 1989
 A:Title: DNA sequence and genetic analysis of the Rhodobacter capsulatus nifX gene
 A:Reference number: JE0029; MUID:89313674; PMID:2747620
 A:Accession: JE0031
 A:Molecule type: DNA
 A:Residues: 1-159 <MOR>
 A:Cross-references: GB:X17433; NID:946085; PIDN:CAA35474.1; PID:946088
 C:Genetics:
 A:Gene: nifX
 C:Keywords: ATP; nitrogen fixation; oxidoreductase

Query Match
 Best Local Similarity 3.2%; Score 7; DB 2; Length 159;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FSSATDQ 73
 Db 60 FSSATDQ 66

RESULT 12
 T34009
 hypothetical protein Y49G5A.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34009
 R:Becker, M.; Tin-Mollam, A.M.; Yeakum, M.
 A:Submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of C. elegans cosmid Y49G5A.
 A:Reference number: Z21457

A:Accession: T34009
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-182 <BEC>
 A:Cross-references: EMBL:AF125968; PDB:1AMD14760.1; GSPDB:GN00023; CESP:Y4965A.1
 A:Experimental source: strain Bristol N2; clone Y4965A
 C:Genetics:
 A:Gene: CESP:Y4965A.1
 A:Map position: 5
 A:introns: 44/3; 82/1; 102/3

Query Match 3.2%; Score 7; DB 2; Length 182;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 SKFCPAG 143
 DB 162 SKFCPAG 168

RESULT 13

AG2720
 riboflavin synthase alpha chain [imported] - Agrobacterium tumefaciens (strain C58, Dupo
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
 C:Accession: AG2720
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tso, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: A2577; PMID:11743193
 A:Accession: AG2720
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-203 <KUR>
 A:Cross-references: GB:AE006688; PDB:1AAL42181.1; PDB:17739571; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: ribc
 A:Map position: circular chromosome
 C:Superfamily: riboflavin synthase alpha chain

Query Match 3.2%; Score 7; DB 2; Length 203;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 DELGCHI 183
 DB 96 DELGCHI 102

RESULT 14

C97502
 riboflavin synthase alpha chain (A236918) [imported] - Agrobacterium tumefaciens (stra
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: C97502
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Muller, N.; Blanchard, M.; Quirillo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:11743194
 A:Accession: C97502
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-204 <KUR>
 A:Cross-references: GB:AB007869; PDB:1AAK86972.1; PDB:15156208; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_2161
 A:Map position: circular chromosome

C:Superfamily: riboflavin synthase alpha chain

Query Match 3.2%; Score 7; DB 2; Length 204;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 DELGCHI 183
 DB 97 DELGCHI 103

RESULT 15

JF0074
 beta-lactamase (EC 3.5.2.6), cefotaxime-hydrolyzing - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 22-Apr-1995 #sequence_revision 03-Aug-1995 #text_change 21-Jul-2000
 C:Accession: JF0074; 141146
 R:ishii, Y.; Ohno, A.; Taguchi, H.; Imajo, S.; Ishiguro, M.; Matsuzawa, H.
 submitted to JIPID, July 1995
 A:Description: Cloning and the sequence analysis of a cefotaxime-hydrolyzing class A
 A:Reference number: JF0074
 A:Accession: JF0074
 A:Molecule type: protein
 A:Residues: 1-291 <ISR>
 R:ishii, Y.; Ohno, A.; Taguchi, H.; Matsuzawa, H.; Yamaguchi, K.
 Antimicrob. Agents Chemother. 39, 2269-2275, 1995
 A:Title: Cloning and sequence analysis of the gene for a cefotaxime-hydrolyzing class
 A:Reference number: 141146; PMID:96109425; PMID:8619581
 A:Accession: 141146
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-89, 'D', 91-291 <RES>
 A:Cross-references: GB:D37830; NID:91435024; PDB:1BAA07082.1; PDB:1037162
 C:Superfamily: beta-lactamase I
 C:Keywords: hydrolase

Query Match 3.2%; Score 7; DB 2; Length 291;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 LFFSAT 71
 DB 19 LFFSAT 25

Search completed: May 15, 2003, 13:25:58
 Job time : 22.2527 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:17:05 ; Search time 6.53218 Seconds

(without alignments)
1390.548 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EELGDCGCHIVTSQDSGTMT.....NGVLSRHGSLSEKRFLETPP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 89080

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	3.7	195	1 ZN90_HUMAN	003938 homo sapien
2	8	3.7	353	1 Z254_HUMAN	075437 homo sapien
3	8	3.7	418	1 Z117_HUMAN	003924 homo sapien
4	8	3.7	439	1 UHPC_ECOLI	P09836 escherichia
5	8	3.7	519	1 TBL3_HUMAN	Q12788 homo sapien
6	8	3.7	535	1 Z257_HUMAN	092251 homo sapien
7	8	3.7	574	1 YET3_HUMAN	092255 homo sapien
8	8	3.7	595	1 ZN85_HUMAN	003923 homo sapien
9	7	3.2	159	1 NITX_RHOCA	P19078 rhodobacter
10	7	3.2	180	1 SYB_APICA	P35589 alypsia cal
11	7	3.2	291	1 BLC2_SALTY	P74841 salmonella
12	7	3.2	291	1 BLC3_SALTY	Q33807 salmonella
13	7	3.2	291	1 BLC5_SALTY	065975 salmonella
14	7	3.2	291	1 BLC6_SALTY	065976 salmonella
15	7	3.2	291	1 BLC7_SALTY	047066 escherichia
16	7	3.2	302	1 BCGH_RHOSH	Q92366 rhodobacter
17	7	3.2	359	1 SIA9_MOUSE	088829 mus musculu
18	7	3.2	384	1 PURT_BACSD	P39771 bacillus su
19	7	3.2	385	1 RECF_MYCTU	059586 mycobacteri
20	7	3.2	419	1 K1C4_XENLA	P05781 xenopus lae
21	7	3.2	494	1 VPE_CITSI	P49043 citrus sine
22	7	3.2	607	1 RECO_ECOLI	P15043 escherichia
23	7	3.2	642	1 YM93_YEAST	P40308 saccharomyc
24	6	2.7	102	1 RPOL_AERPE	Q9yea1 aeropyrum p
25	6	2.7	110	1 KDFE_ERMCH	005527 aerwinia chr
26	6	2.7	122	1 RL18_BUCAL	P57575 buchnera ap
27	6	2.7	126	1 YIGF_ECOLI	P27842 escherichia
28	6	2.7	126	1 YIGF_SALTY	P31139 salmonella
29	6	2.7	137	1 ATPE_SYNP7	P07892 synchococc
30	6	2.7	138	1 ATPE_SYNP1	005375 synchococc
31	6	2.7	140	1 NT7_CYPCA	Q93474 cyprinus ca
32	6	2.7	141	1 V16K_TRVPS	P05075 tobacco rat
33	6	2.7	141	1 V16K_TRVSY	P05077 tobacco rat

34	6	2.7	141	1 V16K_TRVTC	P05076 tobacco rat
35	6	2.7	146	1 LU01_APICA	Q00676 alypsia cal
36	6	2.7	148	1 GVP1_AKASP	Q8YU11 anabaena sp
37	6	2.7	148	1 NR01_CORGL	Q9464 corynebacte
38	6	2.7	151	1 LCT2_HUMAN	Q14960 homo sapien
39	6	2.7	157	1 COAD_HELIP	Q26010 helicobacte
40	6	2.7	160	1 IF52_LYCES	Q9axq5 lycopersico
41	6	2.7	160	1 IF53_SOLTU	P56335 solanum tub
42	6	2.7	160	1 UBCC_SCHPO	Q9p611 schizosacch
43	6	2.7	170	1 OPR3_NEIMC	P10171 neisseria m
44	6	2.7	170	1 Y780_AOUAE	Q66975 aquilex aeo
45	6	2.7	173	1 NHP2_YEAST	P32495 saccharomyc

ALIGNMENTS

RESULT 1
ZN90_HUMAN STANDARD; PRT; 195 AA.
AC 003938;
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 90 (Zinc finger protein HTF9) (Fragment).
GN ZNF90.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91219421; PubMed=2023909;
RA Ballefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA Martial J.A.,
RT "The evolutionarily conserved Kruppel-associated box domain defines a
RT subfamily of eukaryotic multifingered proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M61870; AAA36028.1; -
CC PIR: E39384; E39384.
CC Genew; HGNC:13165; ZNF90.
CC MIM: 603973; -
CC InterPro: IPR001909; KRAB.
CC InterPro: IPR000822; Znf.C2H2.
CC Pfam: PF00096; Zf-C2H2; 1.
CC Pfam: PF01352; KRAB; 1.
CC SMART: SM00349; KRAB; 1.
CC SMART: SM00355; Znf.C2H2; 1.
CC PROSITE: PS00805; KRAB; 1.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
CC PROSITE: PS00157; ZINC_FINGER_C2H2_2; 1.
CC Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
CC Nuclear protein; Repeat.
CC DOMAIN 4
CC ZN_FING 145 167 KRAB.
CC ZN_FING 173 195 C2H2-TYPE (DEGENERATE).
CC ZN_FING 195 195 C2H2-TYPE.
CC NON_TER 195 195
CC SEQUENCE 195 AA; 22562 MW; 0E808A1A62342608 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 195;
 Best Local Similarity 100.0%; Pred. No. 0.93;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||
 DB 50 PDLITCLE 57

RESULT 2

2254_HUMAN STANDARD; PRT; 353 AA.
 AC 075437;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc finger protein 254 (Bone marrow zinc finger 5) (BMZF-5)
 DE (hematopoietic cell derived zinc finger protein 1) (HD-ZNF1).
 GN ZNF254 OR BMZF5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RX MEDLINE=98318631; PubMed=9653160;
 RA Mo M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
 Ha K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
 Wang Y.-X., Chen S.-J., Chen Z.;
 RT "Identification of genes expressed in human CD34(+) hematopoietic
 stem/progenitor cells by expressed sequence tags and efficient full-
 length cDNA cloning.";
 RT Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20054457; PubMed=10585455;
 RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
 Shi S.-L., Zhou J., Han Z.-G., Mao M., Chen S.-J., Yu L., Chen Z.;
 RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
 hematopoietic cells and identification of a novel transregulatory
 domain KRNB.";
 RT J. Biol. Chem. 274:35741-35748(1999).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF054180; AAC39913.1; -
 CC Genew; HGNC:13047; ZNF254.
 CC MIM: 604768; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 4.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZNCFINGER.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00353; Znf_C2H2; 3.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 4.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;

KW Nuclear protein; Repeat.
 FT DOMAIN 13 84
 FT KRAB.
 FT ZINC_FINGERS.
 FT ZN_FING 210 316
 FT C2H2-TYPE (DEGENERATE).
 FT ZN_FING 210 232
 FT C2H2-TYPE (DEGENERATE).
 FT ZN_FING 238 260
 FT C2H2-TYPE.
 FT ZN_FING 266 288
 FT ZN_FING 294 316
 FT C2H2-TYPE.
 SO SEQUENCE 353 AA; 41698 MW; D923CB82D2C5B56A CRC64;

Query Match 3.7%; Score 8; DB 1; Length 353;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||
 DB 59 PDLITCLE 66

RESULT 3

2117_HUMAN STANDARD; PRT; 418 AA.
 AC 003924;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc finger protein 117 (Zinc finger protein HFP9) (Fragment).
 GN ZNF117.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_Taxid=9606;
 RN [1]
 RP SEQUENCE OF 4-418 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=69377476; PubMed=2505992;
 RA Bellefroid E.J., Lecocq P.J., Benhida A., Poncelet D.A.,
 Belayew A., Martial J.A.;
 RT "The human genome contains hundreds of genes coding for finger
 RT proteins of the Kruppel type.";
 RT DNA 8:377-387(1989).
 RN [2]
 RP SEQUENCE OF 1-166 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 Martial J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RT Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M27879; -; NOT ANNOTATED_CDS.
 CC Genew; M61867; AAAS8666.1; -
 DR PIR: B39384; B39384.
 DR HSP: P03001; ITF6.
 DR Genew; HGNC:12897; ZNF117.
 DR MIM: 194624; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 10.
 DR ProDom: PD000003; Znf_C2H2; 5.

DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; ZNF_C2H2; 10.
 DR PROSITE; PS00805; ZINC_FINGER_C2H2_1; 7.
 DR PROSITE; PS00038; ZINC_FINGER_C2H2_2; 9.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 9.
 DR Transcription regulation; zinc-finger; DNA-binding; Metal-binding;
 KW Nuclear protein; Repeat.
 FT NON_TER 1
 FT DOMAIN <1 46 KRAB.
 FT ZN_FING 116 417 ZINC_FINGERS.
 FT ZN_FING 116 138 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 144 166 C2H2-TYPE.
 FT ZN_FING 172 194 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 200 222 C2H2-TYPE.
 FT ZN_FING 228 250 C2H2-TYPE.
 FT ZN_FING 256 278 C2H2-TYPE.
 FT ZN_FING 284 306 C2H2-TYPE.
 FT ZN_FING 312 334 C2H2-TYPE.
 FT ZN_FING 340 362 C2H2-TYPE.
 FT ZN_FING 368 390 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 396 417 C2H2-TYPE.
 FT CONFLICT 75 75 R -> G (IN REF. 2).
 FT CONFLICT 78 79 GY -> RH (IN REF. 2).
 FT CONFLICT 78 79
 SQ SEQUENCE 418 AA; 48941 MW; CDB554CB69EB0B69 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 119 PDLTCL 126
 DB 21 PDLTCL 28
 RESULT 4
 UHPC_ECOLI STANDARD; PRT; 439 AA.
 ID UHPC_ECOLI
 AC P09836; P76728;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Regulatory protein uhpc.
 GN UHPC OR B3667.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87279903; Pubmed=3301805;
 RA Friedrich M.J.; Kadner R.J.;
 RT "Nucleotide sequence of the uhpc region of Escherichia coli.";
 RL J. Bacteriol. 169:3556-3563(1987).
 RN [2]
 RP REVISIONS.
 RA Kadner R.J.;
 RL Submitted (Aug-1990) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92234930; Pubmed=1569007;
 RA Island M.D.; Wei B.-Y.; Kadner R.J.;
 RT "Structure and function of the uhpc genes for the sugar phosphate
 transport system in Escherichia coli and Salmonella typhimurium.";
 RL J. Bacteriol. 174:2754-2762(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=93351513; Pubmed=7686882;
 RA Burland V.D.; Plunkett G. II; Daniels D.L.; Blattner F.R.;
 RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
 genome: organizational symmetry around the origin of replication.";
 RL Genomics 16:551-561(1993).
 RN [5]

RP POSSIBLE ROLE.
 RX MEDLINE=93352407; Pubmed=8349544;
 RA Island M.D.; Kadner R.J.;
 RT "Interplay between the membrane-associated UhpB and UhpC regulatory
 proteins.";
 RL J. Bacteriol. 175:5028-5034(1993).
 CC -1- FUNCTION: UHPC IS REQUIRED FOR UHPB EXPRESSION, IT MAY ACT JOINTLY
 WITH THE UHPB SIGNALING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE SLC37A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M17102; AAA24722.1; ALT_INIT.
 DR EMBL; M89479; AAA24726.1; -.
 DR EMBL; I10328; AAA62019.1; ALT_INIT.
 DR EMBL; AE000444; AAC76690.1; ALT_INIT.
 DR PIR; G41853; KEGCUC.
 DR EcoGene; EG11053; uhpc.
 DR InterPro; IPR000849; G1P transporter.
 DR InterPro; IPR003662; sub transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR TRIGRAMS; TIGR00881; ZAO104; 1.
 DR PROSITE; PS00942; GLPT; 1.
 DR KW Transport; Sugar transport; Transmembrane; Inner membrane;
 KM Complete proteome.
 FT TRANSMEM 67 87 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 163 183 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 245 265 POTENTIAL.
 FT TRANSMEM 290 310 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 350 370 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 411 431 POTENTIAL.
 SQ SEQUENCE 439 AA; 48256 MW; D8097864B7877700 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 439;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 198 LANGVLSR 205
 DB 57 LANGVLSR 64
 RESULT 5
 TBL3_HUMAN STANDARD; PRT; 519 AA.
 ID TBL3_HUMAN
 AC Q12788;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE WD-repeat protein SAZD (Transducin beta-like 3 protein).
 GN TBL3 OR SAZD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94140377; Pubmed=8307582;
 RA Weinstein Saslow D.L.; Germino G.G.; Somlo S.; Reiders S.T.;
 RT "A transducin-like gene maps to the autosomal dominant polycystic
 kidney disease gene region.";
 RL Genomics 18:709-711(1993).
 RN [5]

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U02609; AAA18945.1; -.
DR Genew; HGNC:11587; TBL3.
DR MIM; 605915; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40.11.
DR PRINTS; PD000320; GPROTEINRPT.
DR PRODOM; PD000018; WD40.9.
DR SMART; SM00320; WD40.9.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 7.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
FT REPEAT 1 23 WD 1.
FT REPEAT 35 67 WD 2.
FT REPEAT 79 109 WD 3.
FT REPEAT 260 290 WD 4.
FT REPEAT 305 336 WD 5.
FT REPEAT 363 393 WD 6.
FT REPEAT 447 477 WD 7.
SQ SEQUENCE 519 AA; 56047 MW; EB08B46952499BD CRC64;
Query Match 3.7%; Score 8; DB 1; Length 519;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 65 LFFSSATD 72
Db 53 LFFSSATD 60
RESULT 6
ID 2257_HUMAN STANDARD; PRT; 535 AA.
AC Q9Y2Q1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).
GN ZNF257 OR BMZF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=20054457; PubMed=10585455;
RA Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
RA Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
RT "Molecular cloning of six novel Kruppel-like zinc finger genes from
RT hematopoietic cells and identification of a novel transregulatory
RT domain KRNB".
RL J. Biol. Chem. 274:35741-35748(1999).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF070651; AAD20957.1; -.
DR HSSP; P08047; ISP2.
DR Genew; HGNC:13498; ZNF257.
DR MIM; 606957; -.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR000822; ZnF_C2H2.
DR Pfam; PF00096; ZF_C2H2_10.
DR Pfam; PF01352; KRAB; 1.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; ZnF_C2H2; 9.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZnF_C2H2; 10.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 4 75 KRAB.
FT ZN_FING 173 509 ZINC_FINGERS.
FT ZN_FING 173 195 C2H2-TYPE.
FT ZN_FING 201 223 C2H2-TYPE.
FT ZN_FING 229 251 C2H2-TYPE.
FT ZN_FING 257 282 C2H2-TYPE (DEGENERATE).
FT ZN_FING 288 310 C2H2-TYPE.
FT ZN_FING 316 338 C2H2-TYPE.
FT ZN_FING 344 366 C2H2-TYPE.
FT ZN_FING 372 397 C2H2-TYPE (ATYPICAL).
FT ZN_FING 403 425 C2H2-TYPE.
FT ZN_FING 431 453 C2H2-TYPE.
FT ZN_FING 459 481 C2H2-TYPE.
FT ZN_FING 487 509 C2H2-TYPE.
SQ SEQUENCE 535 AA; 62348 MW; 22DC5B0C4613BC51 CRC64;
Query Match 3.7%; Score 8; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 119 PDLITCLE 126
Db 50 PDLITCLE 57
RESULT 7
ID YE73_HUMAN STANDARD; PRT; 574 AA.
AC Q9P255;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIA1473 (fragment).
GN KIA1473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro".
RL DNA Res. 7:143-150(2000).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITTLE 126
| | | | |
DB 50 PDLITTLE 57

RESULT 9

NIFX_RHOCA
ID NIFX_RHOCA STANDARD; PRT; 159 AA.
AC P19078;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Nifx protein.
GN Nifx.
OS Rhodospirillum rubrum (Rhodospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
CC Rhodospirillum.
CC NCB1_TaxID=1061;
CC [1]

SEQUENCE FROM N.A.
RX MEDLINE=89313674; PubMed=2747620;
RA Moreno-Vivian C., Schmehl M., Masepohl B., Arnold W., Klipp W.;
RT "DNA sequence and genetic analysis of the Rhodospirillum rubrum
Nifx gene region: homology between Nifx and NifB suggests
involvement of Nifx in processing of the iron-molybdenum cofactor."
RL Mol. Gen. Genet. 216:353-363(1989).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE PROCESSING OF THE IRON-
MOLYBDENUM COFACTOR.
CC -1- SIMILARITY: TO NIFB.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: X17433; CAA5474.1; -
DR PIR: J00031; J00031.
DR InterPro: IPR003731; DUF153.
DR Pfam: PF02579; DUF153; 1.
DR Nitrogen fixation.
KW SEQUENCE 159 AA; 17456 MW; 8470D344B6C71005 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FSSARDQ 73
| | | | |
DB 60 FSSARDQ 66

RESULT 10
SYB_SYB_APLICA STANDARD; PRT; 180 AA.

AC P3589;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Synaptobrevin.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidia;
CC Aplysiidae; Aplysia.
CC NCB1_TaxID=6500;
CC [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Ganglion;
RX MEDLINE=94255392; PubMed=8197120;

RA Yamasaki S., Hu Y., Binz T., Kalkuhl A., Kurazono H., Tamura T.,
RA Jahn R., Kandel E., Niemann H.;
RT "Synaptobrevin/vesicle-associated membrane protein (VAMP) of Aplysia
RT californica: structure and proteolysis by tetanus toxin and botulinum
RT neurotoxins type D and F."
RL Proc. Natl. Acad. Sci. U.S.A. 91:4688-4692(1994).
CC -1- FUNCTION: INTRINSIC MEMBRANE PROTEIN OF SMALL SYNAPTIC
CC VESICLES.

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NEURONAL SYNAPTIC
CC VESICLES.

CC -1- SIMILARITY: BELONGS TO THE SYNAPTOSOMAL FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: U00997; AAA19442.1; -
DR InterPro: IPR001388; Synaptobrevin.
DR Pfam: PF00957; synaptobrevin; 1.
DR PRINTS: PR00219; SYNAPTOSOMAL.
DR PRODOM: PD001229; synaptobrevin; 1.
DR PROSITE: PS00417; SYNAPTOSOMAL; 1.
KW Synapse; Transmembrane; Synaptosome; Signal-anchor.
FT DOMAIN 1 84 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 85 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT
FT DOMAIN 105 180 VESICULAR (PROBABLE).
FT SEQUENCE 180 AA; 19746 MW; ADC7780DA74CF347 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VTSQDSG 17
| | | | |
DB 105 VTSQDSG 111

RESULT 11
BLC2_SALTY
ID BLC2_SALTY STANDARD; PRT; 291 AA.

AC P74841;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE beta-lactamase CTX-M-2 precursor (EC 3.5.2.6) (Cefotaxime 2).
GN BLa.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Salmonella.

CC NCB1_TaxID=602;
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=CAS-5;
RX MEDLINE=96431831; PubMed=8834913;

RA Bauernfeind A., Stemplinger I., Jungwirth R., Casellas J.M.;
RT "Sequences of beta-lactamase genes encoding CTX-M-1 (MEN-1) and CTX-M-
RT 2 and relationship of their amino acid sequences with those of other
RT beta-lactamases."
RL Antimicrob. Agents Chemother. 40:509-513(1996).
CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
CC amino acid.

CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X92507; CA63263.1; -.
DR HSSP: Q47066; 1BZA.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
FT SIGNAL 1 28
FT CHAIN 29 291 BY SIMILARITY.
FT ACT_SITE 73 73 BETA-LACTAMASE CTX-M-2.
FT BINDING 237 239 BY SIMILARITY.
SQ SEQUENCE 291 AA; 31377 MW; 82277CD9CDA1E7C0 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSAT 71
Db 19 LFFSSAT 25

RESULT 12
BL66_SALTY
ID BL66_SALTY STANDARD; PRT; 291 AA.
AC 03807;

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-lactamase CTX-M-3 precursor (EC 3.5.2.6) (cefotaxime 3).
GN BLA.
OS Salmonella typhimurium.
OG Plasmid pMS1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;

RN (1)
RP SEQUENCE FROM N.A.
RA Gazouli M.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O - a substituted beta-amino acid.

CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Y14156; CA74573.1; -.
DR HSSP: Q47066; 1BZA.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
FT SIGNAL 1 28
FT CHAIN 29 291 BY SIMILARITY.
FT ACT_SITE 73 73 BETA-LACTAMASE CTX-M-3.
FT BINDING 237 239 BY SIMILARITY.
SQ SEQUENCE 291 AA; 31254 MW; 2E22E251008DF7C6 CRC64;

Query Match 3.2%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSAT 71
Db 19 LFFSSAT 25

RESULT 13

BL66_SALTY
ID BL66_SALTY STANDARD; PRT; 291 AA.
AC 065975;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-lactamase CTX-M-5 precursor (EC 3.5.2.6) (cefotaxime 5).
GN BLA.
OS Salmonella typhimurium.
OG Plasmid pAS30.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;

RN (1)
RP SEQUENCE FROM N.A.
RA Gazouli M., Tzelepi E., Markogiannakis A., Legakis N., Tzouveleakis L.;
RC STRAIN=AS30;
RT "Two novel plasmid-mediated cefotaxime-hydrolyzing beta-lactamases
RT (CTX-M-5 and CTX-M-6) from Salmonella typhimurium."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O - a substituted beta-amino acid.

CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AJ005044; CA06311.1; -.
DR HSSP: Q47066; 1BZA.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
FT SIGNAL 1 28
FT CHAIN 29 291 BY SIMILARITY.
FT ACT_SITE 73 73 BETA-LACTAMASE CTX-M-5.
FT BINDING 237 239 BY SIMILARITY.
SQ SEQUENCE 291 AA; 31251 MW; DCC1C3F3789894E CRC64;

Query Match 3.2%; Score 7; DB 1; Length 291;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 LFFSSAT 71
Db 19 LFFSSAT 25

RESULT 14

BL66_SALTY
ID BL66_SALTY STANDARD; PRT; 291 AA.
AC 065976;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Beta-lactamase CTX-M-6 precursor (EC 3.5.2.6) (cefotaximase 6).
 GN BIA.
 OS *Salmonella typhimurium*.
 CC Plasmid PAS31.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Salmonella*.
 RN NCBI_Taxid=602;
 RP SEQUENCE FROM N.A.
 RC STRAIN=AS31;
 RA Gazouli M., Tzelepi E., Markogiannakis A., Legakis N., Tzouveleakis L.;
 RT "Two novel plasmid-mediated cefotaxime-hydrolyzing beta-lactamases
 (CTX-M-5 and CTX-M-6) from *Salmonella typhimurium*.";
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBD databases.
 CC -1- FUNCTION: HAS CEFOTAXIME-HYDROLYZING ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
 CC amino acid.
 CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ005045; CA06312.1; -
 CC HSSP: Q47066; 1BZA.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
 KW Hydrolyase; Antibiotic resistance; Plasmid; Signal.
 FT SIGNAL 1 28 BY SIMILARITY.
 FT CHAIN 29 291 BETA-LACTAMASE CTX-M-6.
 FT ACT_SITE 73 73 BY SIMILARITY.
 FT BINDING 237 239 SUBSTRATE (BT SIMILARITY).
 SQ SEQUENCE 291 AA; 31206 MW; DC95E4FE7531053 CRC64;
 Query Match 3.2%; Score 7; DB 1; Length 291;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 LFFSSAT 71
 Db 19 LFFSSAT 25
 RESULT 15
 BLTL_ECOLI STANDARD; PRT; 291 AA.
 AC Q47066;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Beta-lactamase Toho-1 precursor (EC 3.5.2.6).
 GN BIA.
 OS *Escherichia coli*.
 CC Plasmid.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 RN NCBI_Taxid=562;
 RP SEQUENCE FROM N.A.
 RC STRAIN=TOH12191;
 RX MEDLINE=96109425; PubMed=8619581;
 RA Ishii Y., Ohno A., Taguchi H., Imajo S., Ishiguro M., Matsuzawa H.;
 RT "Cloning and sequence of the gene encoding a cefotaxime-hydrolyzing
 RT class A beta-lactamase isolated from *Escherichia coli*.";
 RL Antimicrob. Agents Chemother. 39:2269-2275(1995).
 RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT ALA-169.
 RC STRAIN=TOH12191;
 RX MEDLINE=99126537; PubMed=9925786;
 RA Ibuka A., Taguchi A., Ishiguro M., Fushinobu S., Ishii Y.,
 RA Kamitori S., Okuyama K., Yamaguchi K., Kono M., Matsuzawa H.;
 RT "Crystal structure of the E166A mutant of extended-spectrum
 RT beta-lactamase Toho-1 at 1.8 Å resolution".
 RL J. Mol. Biol. 285:2079-2087(1999).
 CC -1- FUNCTION: HAS STRONG CEFOTAXIME-HYDROLYZING ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
 CC amino acid.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D37830; BAA07082.1; -
 CC PDB: 1BZA; 27-APR-99.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
 KW Hydrolyase; Antibiotic resistance; Signal; Plasmid; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 291 BETA-LACTAMASE TOHO-1.
 FT ACT_SITE 73 73 SUBSTRATE.
 FT BINDING 237 239
 SQ SEQUENCE 291 AA; 31447 MW; 83FC0CDD9CD41E7C0 CRC64;
 Query Match 3.2%; Score 7; DB 1; Length 291;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 LFFSSAT 71
 Db 19 LFFSSAT 25
 Search completed: May 15, 2003, 13:23:08
 Job time : 9.53218 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:18:56 ; Search time 18.9089 Seconds
(without alignments)
2386.400 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EELGDCGHIVTSQDSGTMT.....NCVLSRHSLSEKRELFPTTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 504360

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

SPRKBML_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	100.0	503	11	09D4J3 mus musculus
2	201	91.8	460	11	09D696 mus musculus
3	145	66.2	432	11	08R327 mus musculus
4	8	3.7	195	17	08R327 mus musculus
5	8	3.7	216	4	096H40 homo sapien
6	8	3.7	336	10	09MAA9 Oryza sativa
7	8	3.7	364	4	09H8G1 homo sapien
8	8	3.7	440	16	08X543 Mus musculus
9	8	3.7	470	4	09B2E6 Mus musculus
10	8	3.7	519	4	09B604 Mus musculus
11	8	3.7	524	4	0965C4 Mus musculus
12	8	3.7	536	4	096B85 Mus musculus
13	8	3.7	532	5	08T1H2 Mus musculus
14	8	3.7	532	5	08T1H2 Mus musculus
15	8	3.7	532	5	08T1H2 Mus musculus
16	8	3.7	532	5	08T1H2 Mus musculus

17	7	3.2	194	11	091YF2 Mus musculus
18	7	3.2	204	16	08G772 Mus musculus
19	7	3.2	205	16	08G381 Mus musculus
20	7	3.2	206	16	0920U4 Mus musculus
21	7	3.2	251	7	031296 Mus musculus
22	7	3.2	272	16	08R2M5 Mus musculus
23	7	3.2	276	2	08VQ11 Mus musculus
24	7	3.2	291	2	0936V5 Mus musculus
25	7	3.2	291	2	09RLX4 Mus musculus
26	7	3.2	291	2	08VN25 Mus musculus
27	7	3.2	291	2	08VN24 Mus musculus
28	7	3.2	291	2	08VN23 Mus musculus
29	7	3.2	291	2	08VN22 Mus musculus
30	7	3.2	291	2	08VKT5 Mus musculus
31	7	3.2	291	2	034177 Mus musculus
32	7	3.2	294	2	0931I6 Mus musculus
33	7	3.2	300	5	08S029 Mus musculus
34	7	3.2	314	17	097287 Mus musculus
35	7	3.2	320	5	09XMB9 Mus musculus
36	7	3.2	332	13	091333 Mus musculus
37	7	3.2	350	4	09NH83 Mus musculus
38	7	3.2	350	4	09NH83 Mus musculus
39	7	3.2	350	11	09C2V9 Mus musculus
40	7	3.2	350	11	09D771 Mus musculus
41	7	3.2	359	11	09QWF8 Mus musculus
42	7	3.2	359	11	088829 Mus musculus
43	7	3.2	368	5	09G566 Mus musculus
44	7	3.2	381	10	09FNP5 Mus musculus
45	7	3.2	385	3	08X058 Mus musculus

ALIGNMENTS

RESULT 1	ID	09D4J3	PRELIMINARY:	PRT:	503 AA.
AC	09D4J3	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	463143K1R1K protein.				
GN	463143K1R1K				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,				
RA	Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Glisic C., King B., Kociba H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barish G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Williams L.,				
RA	Yuzhaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,				
RA	Hayashizaki Y.,				
RT	Functional annotation of a full-length mouse cDNA collection.;				
RL	Nature 409:685-690(2001).				
CC	-1- SIMILARITY: CONTAINS 1 CDB DOMAIN.				
DR	EMBL: AK016485; BAB30265.1; -.				

DR MGD; MGI:1913936; 4631413K11R1K.
 DR InterPro: IPR000859; CUB_dom.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SQ SEQUENCE 503 AA; 54547 MW; FEB121E845CA06B CRC64;

Query Match 100.0%; Score 219; DB 11; Length 503;
 Best Local Similarity 100.0%; Pred. No. 1.2e-224;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELDGGCHITTSQDSGMTSKNPGTPTVCEKITTVKGRKLLRLGDLNIESKTC 60
 |||||||
 DB 26 EELDGGCHITTSQDSGMTSKNPGTPTVCEKITTVKGRKLLRLGDLNIESKTC 85
 QY 61 ASDYLFSATDQGYPCGSAVPELRLNSNEVTLFKSGSHISGRGFLTYASSDHPD 120
 |||||||
 DB 86 ASDYLFSATDQGYPCGSAVPELRLNSNEVTLFKSGSHISGRGFLTYASSDHPD 145
 QY 121 LITCLERSHYEERKSKFCPCAGCDIADISGNTKDGYPDTSLLCKAAIHAGITDELG 180
 |||||||
 DB 146 LITCLERSHYEERKSKFCPCAGCDIADISGNTKDGYPDTSLLCKAAIHAGITDELG 205
 QY 181 GHINLQSKGSHYEGGLANGVLSRHGSLSEKRFLETPP 219
 |||||||
 DB 206 GHINLQSKGSHYEGGLANGVLSRHGSLSEKRFLETPP 244

RESULT 2

Q9D696 PRELIMINARY; PRT; 460 AA.

AC Q9D696;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE 4631413K11R1K protein.
 GN 4631413K11R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 Izawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 Schirral L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Orido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
 Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK014521; BAB29409.1; -
 DR MGD; MGI:1913936; 4631413K11R1K.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.

DR PROSITE: PS01180; CUB; 1.
 SQ SEQUENCE 460 AA; 50334 MW; 3CF3356F70A88B93 CRC64;

Query Match 91.8%; Score 201; DB 11; Length 460;
 Best Local Similarity 100.0%; Pred. No. 1.8e-205;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 MTSKNPPTVPTVCEKITTVKGRKLLRLGDLNIESKTCASDYLFSATDQGYPC 78
 |||||||
 DB 1 MTSKNPPTVPTVCEKITTVKGRKLLRLGDLNIESKTCASDYLFSATDQGYPC 60
 QY 79 GSAVPELRLNSNEVTLFKSGSHISGRGFLTYASSDHPDITCLERSHYEERKSK 138
 |||||||
 DB 61 GSAVPELRLNSNEVTLFKSGSHISGRGFLTYASSDHPDITCLERSHYEERKSK 120
 QY 139 FCPACRDIADISGNTKDGYPDTSLLCKAAIHAGITDELGHINLQSKGSHYEGGL 198
 |||||||
 DB 121 FCPACRDIADISGNTKDGYPDTSLLCKAAIHAGITDELGHINLQSKGSHYEGGL 180
 QY 199 ANGVLSRHGSLSEKRFLETPP 219
 |||||||
 DB 181 ANGVLSRHGSLSEKRFLETPP 201

RESULT 3

Q8R327 PRELIMINARY; PRT; 432 AA.

AC Q8R327;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Similar to RIKEN cDNA 4631413K11 gene.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC026771; AAH26771.1; -
 SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 66.2%; Score 145; DB 11; Length 432;
 Best Local Similarity 100.0%; Pred. No. 8.5e-146;
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 GPGGSAVPELRLNSNEVTLFKSGSHISGRGFLTYASSDHPDITCLERSHYEE 134
 |||||||
 DB 29 GPGGSAVPELRLNSNEVTLFKSGSHISGRGFLTYASSDHPDITCLERSHYEE 88
 QY 135 KYSKFCPCAGCDIADISGNTKDGYPDTSLLCKAAIHAGITDELGHINLQSKGSHY 194
 |||||||
 DB 89 KYSKFCPCAGCDIADISGNTKDGYPDTSLLCKAAIHAGITDELGHINLQSKGSHY 148
 QY 195 EGLANGVLSRHGSLSEKRFLETPP 219
 |||||||
 DB 149 EGLANGVLSRHGSLSEKRFLETPP 173

RESULT 4

Q8RTJ3 PRELIMINARY; PRT; 195 AA.

AC Q8RTJ3;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein MA3694.
 GN MA3694.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Maylor J., Stange-Thomann N., DeRellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Pittrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanocaldococcus jannaschii reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011078; AM07049.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 22019 MW; 909E97DB7359DC67 CRC64;

Query Match
Best Local Similarity 3.7%; Score 8; DB 17; Length 195;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 AGITDEL 179
Db 98 AGITDEL 105

RESULT 5
ID 096H40 PRELIMINARY; PRT; 216 AA.
AC 096H40;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE Similar to DNA-binding protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008936; AA08936.1;
DR InterPro; IPR001909; KRAB.
DR Pfam; PF01352; KRAB.
DR Pfam; PF01352; KRAB.
DR PROSITE; PSS0805; KRAB.
DR PROSITE; PSS0157; ZINC_FINGER_C2H2_2; 2.
DR DNA-binding; Zinc-finger.
SQ SEQUENCE 216 AA; 25227 MW; B5C6CF182AFCCE2 CRC64;

Query Match
Best Local Similarity 3.7%; Score 8; DB 4; Length 216;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PDLITCLE 126
Db 50 PDLITCLE 57

RESULT 6
ID 09MAA9 PRELIMINARY; PRT; 336 AA.
AC 09MAA9;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

```

```

DE T12H1.6 protein (Hypothetical 37.7 kDa protein).
GN T12H1.6 OR AT3G05100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Beilto M., Creasy T.H., Haas B., Wu D.,
RA Rongning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T12H1 genomic sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene At3g05100 (GI:15229893)."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009177; AA27016.1;
DR EMBL; AY070370; AA49868.1;
DR EMBL; AY091328; AA414267.1;
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein.
SQ SEQUENCE 336 AA; 37719 MW; 6FA910F7B4B2FA85 CRC64;

Query Match
Best Local Similarity 3.7%; Score 8; DB 10; Length 336;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 KGISHYEG 196
Db 121 KGISHYEG 128

RESULT 7
ID 09H8G1 PRELIMINARY; PRT; 364 AA.
AC 09H8G1;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE CDNA FLJ13659 fts, clone PLACE1011576, moderately similar to human
DE kruppel related zinc finger protein (HRF1) mRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Tsogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sato H., Sugano S., Shitatori A., Sudo H.,

```

RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takehashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Saito K.,
 RA Yamamoto J., Wakematsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Nishimura K., Iwayanagi T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK023721; BAB14656.1; -
 DR InterPro: IPR001909; KRA6.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam; PF01352; KRA6; 1.
 DR Pfam; PF00096; Zf-C2H2; 3.
 DR ProDom; PD000003; Znf_C2H2; 2.
 DR SMART; SM00349; KRA6; 1.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS50805; KRA6; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SO SEQUENCE 364 AA; 42472 MW; 5043DA4C575842B0 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 364;
 Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||

DB 82 PDLITCLE 89

RESULT 8

ID 08X543 PRELIMINARY; PRT; 440 AA.

AC 08X543;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Regulator of Unp.

GN UNP OR Z5157 OR ECS4604.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

NCBI_TaxID=83334;

RA [1]

SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RA MEDLINE=21074935; PubMed=11206551;

RA Pena N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Grolbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RL Nature 409:529-533(2001).

RA [2]

SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RA MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

RA Kunara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

DR EMBL; AE005598; AAC58866.1; -

DR EMBL; AP002366; BAB38027.1; -

DR InterPro: IPR000849; G1PT_transporter.

DR InterPro: IPR003662; sub_transporter.

DR Pfam; PF00083; sugar_tr; 1.

DR TIGRfams: TIGR00881; 2A0104; 1.

DR PROSITE; PS00942; GLPT; 1.

KW Complete proteome.
 SO SEQUENCE 440 AA; 48345 MW; C3442BAC11132448 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 440;
 Best Local Similarity 100.0%; Pred. No. 10;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 198 LANGVLSR 205
 |||||

DB 58 LANGVLSR 65

RESULT 9

ID 09BZ6 PRELIMINARY; PRT; 470 AA.

AC 09BZ6;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE KRA6 zinc finger protein HZF26 (Fragment).

GN HZF26.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

RA [1]

SEQUENCE FROM N.A.

RC MEDLINE=95169271; PubMed=7865130;

RA Abrik M., Aveskogh M., Hellman L.;

RT "Isolation of cDNA clones for 42 different Kruppel-related zinc finger

proteins expressed in the human monoblast cell line U-937.";

RL DNA Cell Biol. 14:125-136(1995).

RA [2]

SEQUENCE FROM N.A.

RC Abrik M., Aveskogh M., Hellman L.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; AF325191; AAK01422.1; -

DR HSP; P08046; 1AIG.

DR InterPro: IPR001909; KRA6.

DR InterPro: IPR000822; Znf_C2H2.

DR Pfam; PF01352; KRA6; 1.

DR Pfam; PF00096; Zf-C2H2; 10.

DR PRINTS; PR00048; ZINC_FINGER.

DR SMART; SM00349; KRA6; 1.

DR SMART; SM00355; Znf_C2H2; 9.

DR PROSITE; PS50805; KRA6; 1.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.

DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 11.

RT DNA-binding; Metal-binding; Nuclear protein; zinc-finger.

FW NON_TER 470

SO SEQUENCE 470 AA; 54961 MW; 997CB1C95D37E983 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 470;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||

DB 50 PDLITCLE 57

RESULT 10

ID 096M04 PRELIMINARY; PRT; 519 AA.

AC 096M04;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE cDNA FLJ32933 fls, clone TEST12007466, moderately similar to zinc

finger protein 91.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Houta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Nagasuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isoigal T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK057495; BAB71510.1;
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf_C2H2; 12.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_8.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
 KW DNA-binding; zinc-finger.
 SQ SEQUENCE 519 AA; 60350 MW; 7EAD17E53EC7732C CRC64;

Query Match 3.7%; Score 8; DB 4; Length 519;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||

DB 59 PDLITCLE 66

RESULT 11
 ID 096JUC4 PRELIMINARY; PRT; 524 AA.
 AC 096JUC4;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE KRAB zinc finger protein.
 GN KR19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RX MEDLINE-21303228; PubMed-11410164;
 RA Mark C., Looman C., Ahrink M., Hellman L.;
 RT "Molecular cloning and preliminary functional analysis of two novel
 RT human KRAB zinc finger proteins, Hkr18 and Hkr19."
 RL DNA Cell Biol. 20:275-286(2001).
 DR EMBL: AF277624; AAK61307.1;
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf_C2H2; 12.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_11.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
 KW DNA-binding; zinc-finger.
 SQ SEQUENCE 524 AA; 60596 MW; F1AD4929DC67D105 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 524;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||

DB 62 PDLITCLE 69

RESULT 12
 ID 096BBS PRELIMINARY; PRT; 536 AA.
 AC 096BBS;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical 61.9 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-B-CELL;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC015765; AAH15765.1;
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; zf_C2H2; 13.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_13.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 13.
 KW DNA-binding; Hypothetical protein; zinc-finger.
 SQ SEQUENCE 536 AA; 61932 MW; 3DC80FAE53F8C00 CRC64;

Query Match 3.7%; Score 8; DB 4; Length 536;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PDLITCLE 126
 |||||||

DB 50 PDLITCLE 57

RESULT 13
 ID 087IH2 PRELIMINARY; PRT; 132 AA.
 AC 087IH2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ATMLSS RASGEF.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachbat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tegenfeldt B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC116548; AAL9317.1;
 SQ SEQUENCE 132 AA; 14408 MW; 9AD15CB58D7666F8 CRC64;

Query Match 3.2%; Score 7; DB 5; Length 132;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 FSSATDQ 73
 |||||||

DB 57 FSSATDQ 63

RESULT 14
 O9PPN3

ID QPPN3 PRELIMINARY; PRT: 135 AA.
 AC QPPN3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative ATP /GTP-binding protein.
 GN C10668.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 CC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jørgensen K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL: AL139076; CAB72945.1;
 DR InterPro: IPR003442; UPP0079.
 DR Pfam: PF02367; UPP0079; 1.
 DR TIGRfams: TIGR00150; UPP0079; 1.
 KW Complete proteome.
 SQ SEQUENCE 135 AA; 15597 MW; 0E50BF8FDCD34E69 CRC64;

Query Match 3.28; Score 7; DB 16; Length 135;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 EGLANG 201
 |||||
 Db 80 EGLANG 86

RESULT 15

Q8TKH7 PRELIMINARY; PRT: 156 AA.
 ID Q8TKH7;
 AC Q8TKH7;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein MA3428.
 GN MA3428.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels S., Smirnov S., Almoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearlano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL: AE011048; AAM06795.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 156 AA; 17812 MW; 184878A5AF292852 CRC64;

Query Match 3.28; Score 7; DB 17; Length 156;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 80 SWAYPKE 86
 |||||
 Db 59 SWAYPKE 65

Search completed: May 15, 2003, 13:24:11
 Job time : 21.9089 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:20:51 ; Search time 13.0644 Seconds
(without alignments)
493.221 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EELGDCGHITSDSGTWT.....NGVLSRHGSLSEKRLFTTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 58762

Minimum DB seq length: 100
Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/CCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.2	334	4 US-09-463-702A-38	Sequence 38, Appl
2	7	3.2	370	2 US-08-559-303B-77	Sequence 77, Appl
3	7	3.2	370	4 US-09-175-828-77	Sequence 77, Appl
4	7	3.2	414	4 US-09-334-601-4	Sequence 4, Appl
5	7	3.2	524	4 US-09-186-276B-56	Sequence 56, Appl
6	7	3.2	524	4 US-08-842-445-56	Sequence 56, Appl
7	7	3.2	524	4 US-09-186-188B-56	Sequence 56, Appl
8	7	3.2	607	3 US-08-781-891-75	Sequence 75, Appl
9	6	2.7	103	2 US-08-796-414B-3	Sequence 3, Appl
10	6	2.7	106	1 US-08-241-853-13	Sequence 13, Appl
11	6	2.7	106	2 US-08-850-917-13	Sequence 13, Appl
12	6	2.7	114	2 US-08-887-352B-11	Sequence 11, Appl
13	6	2.7	114	2 US-08-887-352B-12	Sequence 12, Appl
14	6	2.7	114	4 US-09-109-207C-11	Sequence 11, Appl
15	6	2.7	114	4 US-09-109-207C-12	Sequence 12, Appl
16	6	2.7	114	4 US-09-296-005-11	Sequence 11, Appl
17	6	2.7	114	4 US-09-296-005-12	Sequence 12, Appl
18	6	2.7	121	2 US-08-887-352B-2	Sequence 2, Appl
19	6	2.7	121	2 US-08-887-352B-3	Sequence 3, Appl
20	6	2.7	121	4 US-09-109-207C-2	Sequence 2, Appl
21	6	2.7	121	4 US-09-109-207C-3	Sequence 3, Appl
22	6	2.7	121	4 US-09-296-005-2	Sequence 2, Appl
23	6	2.7	121	4 US-09-296-005-3	Sequence 3, Appl
24	6	2.7	123	2 US-08-626-685A-10	Sequence 10, Appl
25	6	2.7	134	4 US-08-466-151-3	Sequence 3, Appl
26	6	2.7	134	4 US-08-466-153-3	Sequence 3, Appl
27	6	2.7	139	1 US-07-994-469A-10	Sequence 10, Appl

28	6	2.7	145	1 US-07-994-469A-9	Sequence 9, Appl
29	6	2.7	147	1 US-07-668-517-1	Sequence 1, Appl
30	6	2.7	147	4 US-09-105-343A-9	Sequence 9, Appl
31	6	2.7	148	1 US-07-668-517-2	Sequence 2, Appl
32	6	2.7	148	1 US-07-668-517-15	Sequence 15, Appl
33	6	2.7	149	1 US-07-668-517-3	Sequence 3, Appl
34	6	2.7	149	1 US-07-668-517-16	Sequence 16, Appl
35	6	2.7	150	1 US-07-668-517-4	Sequence 4, Appl
36	6	2.7	150	1 US-07-668-517-5	Sequence 5, Appl
37	6	2.7	150	1 US-07-668-517-6	Sequence 6, Appl
38	6	2.7	150	1 US-07-668-517-8	Sequence 8, Appl
39	6	2.7	150	1 US-07-668-517-9	Sequence 9, Appl
40	6	2.7	150	1 US-07-668-517-10	Sequence 10, Appl
41	6	2.7	150	1 US-07-668-517-11	Sequence 11, Appl
42	6	2.7	150	1 US-07-668-517-12	Sequence 12, Appl
43	6	2.7	150	1 US-07-668-517-13	Sequence 13, Appl
44	6	2.7	150	1 US-07-668-517-14	Sequence 14, Appl
45	6	2.7	150	1 US-07-668-517-17	Sequence 17, Appl

ALIGNMENTS

```
RESULT 1
US-09-463-702A-38
; Sequence 38, Application US/09463702A
; Patent No. 6335435
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO, AKIRO
; APPLICANT: KITAO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECD4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/463,702A
; CURRENT FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-463-702A-38
;
; Query Match
; Best Local Similarity 100.0%; Pred. No. 47;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; DB 186 LOSKIS 192
; 236 LOSKIS 242
;
; RESULT 2
US-08-559-303B-77
; Sequence 77, Application US/08559303B
; Patent No. 5824501
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
```

COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-559-303B-77

Query Match 3.2%; Score 7; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 LOSKGIS 192
DB 241 LOSKGIS 247

RESULT 3
US-09-175-828-77
Sequence 77, Application US/09175828
Patent No. 6221543
GENERAL INFORMATION:
APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
APPLICANT: GRODEN
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/175,828
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 370
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-09-175-828-77

Query Match 3.2%; Score 7; DB 4; Length 370;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 LOSKGIS 192
DB 241 LOSKGIS 247

RESULT 4
US-09-334-601-4
Sequence 4, Application US/09334601
Patent No. 6280989
GENERAL INFORMATION:
APPLICANT: Kapitonov, Dmitri
APPLICANT: Yu, Robert
TITLE OF INVENTION: NOVEL STALYLTTRANSFERRASES
FILE REFERENCE: YCUIP-6
CURRENT APPLICATION NUMBER: US/09/334,601
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 414
TYPE: PRT
ORGANISM: Murinae gen. sp.
US-09-334-601-4

Query Match 3.2%; Score 7; DB 4; Length 414;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 VYVLFKS 100
DB 261 VYVLFKS 267

RESULT 5
US-09-186-276B-56
Sequence 56, Application US/09186276B
Patent No. 6388173
GENERAL INFORMATION:
APPLICANT: Bentley, Philip
APPLICANT: DiLaurenzio, Laura
APPLICANT: Wyszocka-Dillier, Joanna
APPLICANT: Malamy, Jocelyn E.
APPLICANT: Pysh, Leonard
APPLICANT: Helariutta, Yrjo
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof

FILE REFERENCE: 5914-075-999
CURRENT APPLICATION NUMBER: US/09/186,276B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 56
LENGTH: 524
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(524)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-186-276B-56

Query Match 3.2%; Score 7; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
DB 335 LRLGDLN 341

RESULT 6
US-08-842-445-56
Sequence 56, Application US/08842445A
Patent No. 6441270
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
FILE REFERENCE: 5914-056-999
CURRENT APPLICATION NUMBER: US/08/842,445A
CURRENT FILING DATE: 1997-04-24
EARLIER APPLICATION NUMBER: 08/638,617
EARLIER FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 56
LENGTH: 524
TYPE: PRT
ORGANISM: Plant
US-08-842-445-56

Query Match 3.2%; Score 7; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
DB 335 LRLGDLN 341

RESULT 7
US-09-186-188B-56
Sequence 56, Application US/09186188B
Patent No. 6455672
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
FILE REFERENCE: 5914-074-999
CURRENT APPLICATION NUMBER: US/09/186,188B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26

NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 56
LENGTH: 524
TYPE: PRT
ORGANISM: Plant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(524)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-186-188B-56

Query Match 3.2%; Score 7; DB 4; Length 524;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
DB 335 LRLGDLN 341

RESULT 8
US-08-781-891-75
Sequence 75, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620, Leburg, Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052,419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 607 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-75

Query Match 3.2%; Score 7; DB 3; Length 607;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 LOSKGS 192
DB 255 LOSKGS 261

RESULT 9
US-08-796-414B-3
; Sequence 3, Application US/08796414B
; Patent No. 5876987
; GENERAL INFORMATION:
; APPLICANT: Wendy C. Champness, Paul Brian
; APPLICANT: and Todd B. Anderson
; TITLE OF INVENTION: METHOD, DNA AND BACTERIA
; TITLE OF INVENTION: FOR HYPERPRODUCTION OF AN
; TITLE OF INVENTION: ANTIBIOTIC DUE TO
; TITLE OF INVENTION: DISRUPTION OF AN ADSA
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch,
; COMPUTER: Acer
; OPERATING SYSTEM: MS-DOS 5.00
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796, 414B
; FILING DATE: February 6, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-297
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 5876987e
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Amino Acid
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces coelicolor
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: unicellular microorganisms
; CELL LINE:
; ORGANELLER:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; FEATURE:
; NAME/KEY: D7
; LOCATION:
; IDENTIFICATION METHOD: deduced
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:

ISSUE:
PAGES:
DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-796-414B-3

Query Match 2.7%; Score 6; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 155 TKDGYR 160
DB 89 TKDGYR 94

RESULT 10
US-08-241-853-13
; Sequence 13, Application US/08241853
; Patent No. 5693488
; GENERAL INFORMATION:
; APPLICANT: Fang, Kathy S.
; APPLICANT: Hanafusa, Hidesaburo
; TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/241,853
; FILING DATE: 12-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-241-853-13
Query Match 2.7%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 35 EKITYV 40
DB 84 EKITYV 89

RESULT 11
US-08-850-917-13
; Sequence 13, Application US/08850917
; Patent No. 5854045

GENERAL INFORMATION:
APPLICANT: Fang, Kathy S.
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE
TITLE OF INVENTION: AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,917
FILING DATE: 02-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/241,853
FILING DATE: 12-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-850-917-13

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 106;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 EXITIV 40
|||||

DB 84 EXITIV 89

RESULT 12
US-08-887-352B-11
Sequence 11, Application US/08887352B
Patent No. 5994511
GENERAL INFORMATION:
APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
TITLE OF INVENTION: Improving Polypeptides
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,352B
FILING DATE: 03-Jul-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1123
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-887-352B-11

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 114;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
|||||

DB 98 RGSHPF 103

RESULT 13
US-08-887-352B-12
Sequence 12, Application US/08887352B
Patent No. 5994511
GENERAL INFORMATION:
APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
TITLE OF INVENTION: Improving Polypeptides
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,352B
FILING DATE: 03-Jul-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1123
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-887-352B-12

Query Match
Best Local Similarity 100.0%; Score 6; DB 2; Length 114;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
|||||

Db 98 RGSHPF 103

```
RESULT 14
US-09-109-207C-11
; Sequence 11, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; CURRENT FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/051,554
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 11
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-114
; OTHER INFORMATION: Heavy chain sequence derived from MAE11
US-09-109-207C-11
```

```
Query Match
2.7%; Score 6; DB 4; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 RGSHPF 132
Db 98 RGSHPF 103
```

```
RESULT 15
US-09-109-207C-12
; Sequence 12, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; CURRENT FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/051,554
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 44
; SEQ ID NO 12
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-114
; OTHER INFORMATION: Heavy chain sequence derived from MAE11
US-09-109-207C-12
```

```
Query Match
2.7%; Score 6; DB 4; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 127 RGSHPF 132
Db 98 RGSHPF 103

Search completed: May 15, 2003, 13:24:51
Job time : 14.0644 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:22:46 ; Search time 11.6892 Seconds
(without alignments)
1807.034 Million cell updates/sec

Title: US-10-003-132-4_COPY_26_244

Perfect score: 219
Sequence: 1 EELGDCGHIVTSQDSGTMT.....NCVLSRHGSLSEKRELFPTTP 219

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 362588 seqs, 96450795 residues

Word size : 0

Total number of hits satisfying chosen parameters: 204442

Minimum DB seq length: 100

Maximum DB seq length: 700

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	503	US-10-003-132-4	Sequence 4, Appl1
2	148	67.6	503	US-09-823-038A-51	Sequence 51, Appl1
3	100	45.7	421	US-10-003-132-6	Sequence 6, Appl1
4	28	12.8	458	US-09-759-130B-76	Sequence 76, Appl1
5	28	12.8	681	US-09-759-130B-75	Sequence 75, Appl1
6	7	3.2	359	US-10-309-389-2	Sequence 2, Appl1
7	7	3.2	370	US-09-753-143-77	Sequence 77, Appl1
8	7	3.2	385	US-09-712-363-148	Sequence 148, Appl1
9	7	3.2	524	US-10-253-007-56	Sequence 56, Appl1
10	7	3.2	604	US-10-045-170A-1	Sequence 1, Appl1
11	7	3.2	610	US-09-815-242-10414	Sequence 10414, A
12	6	2.7	100	US-09-864-761-47749	Sequence 47749, A
13	6	2.7	104	US-09-911-777-5	Sequence 5, Appl1
14	6	2.7	107	US-09-864-761-47562	Sequence 47562, A
15	6	2.7	108	US-09-910-150-21	Sequence 21, Appl1
16	6	2.7	110	US-09-864-761-42215	Sequence 42215, A
17	6	2.7	112	US-09-910-150-19	Sequence 19, Appl1
18	6	2.7	114	US-09-920-171-11	Sequence 11, Appl1
19	6	2.7	114	US-09-920-171-12	Sequence 12, Appl1

20	6	2.7	118	10	US-09-910-150-23	Sequence 23, Appl1
21	6	2.7	121	10	US-09-920-171-2	Sequence 2, Appl1
22	6	2.7	121	10	US-09-920-171-3	Sequence 3, Appl1
23	6	2.7	123	1	US-08-899-112-10	Sequence 10, Appl1
24	6	2.7	123	1	US-09-796-692-2374	Sequence 2374, Ap
25	6	2.7	123	9	US-10-040-862-2374	Sequence 2374, Ap
26	6	2.7	134	9	US-09-925-179-3	Sequence 3, Appl1
27	6	2.7	134	10	US-09-802-077-3	Sequence 3, Appl1
28	6	2.7	134	10	US-09-802-096-3	Sequence 3, Appl1
29	6	2.7	134	10	US-09-815-242-10513	Sequence 10513, A
30	6	2.7	148	9	US-09-738-626-6992	Sequence 6992, Ap
31	6	2.7	148	9	US-09-925-299-800	Sequence 800, App
32	6	2.7	148	10	US-09-925-299-800	Sequence 800, App
33	6	2.7	150	10	US-09-877-156-25	Sequence 25, Appl1
34	6	2.7	150	12	US-10-116-378-28	Sequence 28, Appl1
35	6	2.7	152	9	US-09-779-050A-24	Sequence 24, Appl1
36	6	2.7	157	9	US-09-803-327A-7	Sequence 7, Appl1
37	6	2.7	157	9	US-10-043-432-1	Sequence 1, Appl1
38	6	2.7	157	10	US-09-756-301A-1	Sequence 1, Appl1
39	6	2.7	157	10	US-09-927-703-1	Sequence 1, Appl1
40	6	2.7	157	10	US-09-854-280-19	Sequence 19, Appl1
41	6	2.7	157	10	US-09-934-465-13	Sequence 13, Appl1
42	6	2.7	157	10	US-09-766-535A-1	Sequence 1, Appl1
43	6	2.7	157	10	US-09-854-208-19	Sequence 19, Appl1
44	6	2.7	157	10	US-09-756-161A-1	Sequence 1, Appl1
45	6	2.7	157	12	US-10-010-229-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-10-003-132-4
Sequence 4, Application US/10003132
Publication No. US20020192750A1
GENERAL INFORMATION:
APPLICANT: Fox, Brian A.
APPLICANT: Gao, Zeren
APPLICANT: Shoemaker, Kimberly E.
TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
FILE REFERENCE: 00-62
CURRENT APPLICATION NUMBER: US/10/003,132
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/249,004
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 503
TYPE: PRT
ORGANISM: Mus musculus
US-10-003-132-4

Query Match 100.0%; Score 219; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.3e-204;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELGDCGHIVTSQDSGTMTSKNYPGTYPTVCETITPKKRLILRGDINISKTC 60
DB 26 EELGDCGHIVTSQDSGTMTSKNYPGTYPTVCETITPKKRLILRGDINISKTC 85
QY 61 ASDYLLFSATPDQGYPCGSAWPAKELRLNSNVTYLFKSGHSISGRGFLITAASSDHPD 120
DB 86 ASDYLLFSATPDQGYPCGSAWPAKELRLNSNVTYLFKSGHSISGRGFLITAASSDHPD 145
QY 121 LITCLERGSYFPEKYSKPCPCACRIDIGISGNTFGYRDTSLCKAAIHAGITIDELG 180
DB 146 LITCLERGSYFPEKYSKPCPCACRIDIGISGNTFGYRDTSLCKAAIHAGITIDELG 205
QY 181 GHTNLQSKGISHYEGILANGVLSRHGSLSEKRELFPTTP 219
DB 206 GHTNLQSKGISHYEGILANGVLSRHGSLSEKRELFPTTP 244

RESULT 2
US-09-823-038A-51
; Sequence 51, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Ormstedt, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murlson, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823.038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mouse
US-09-823-038A-51

Query Match 67.6%; Score 148; DB 10; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.3e-135;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EELDGGCHITSDSGMTSKNTPGTYPNATYCKEITTVKRGKLLIRLGDNLIESKTC 60
DB 26 EELDGGCHITSDSGMTSKNTPGTYPNATYCKEITTVKRGKLLIRLGDNLIESKTC 85
QY 61 ASDYLFESSATDQXGPGYGSNAVPEKELRLNSNEVTLFKSGSHISGRFLTYASSDHPD 120
DB 86 ASDYLFESSATDQXGPGYGSNAVPEKELRLNSNEVTLFKSGSHISGRFLTYASSDHPD 145
QY 121 LITCLERGSHTFEKYSKFCPCAGCDIA 148
DB 146 LITCLERGSHTFEKYSKFCPCAGCDIA 173

RESULT 3
US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003.132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6

Query Match 45.7%; Score 100; DB 9; Length 458;
Best Local Similarity 100.0%; Pred. No. 5.2e-89;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 DLITCLERGSHTFEKYSKFCPCAGCDIAGDISGNTKDGVRTSLCKAAIHAGITDEL 179
DB 100 DLITCLERGSHTFEKYSKFCPCAGCDIAGDISGNTKDGVRTSLCKAAIHAGITDEL 159
QY 180 GGHINLQSKGISHYEGILANGVLSRHGSLSEKRFLLFTTP 219

DB 160 GGHINLQSKGISHYEGILANGVLSRHGSLSEKRFLLFTTP 199

RESULT 4
US-09-759-130B-76
; Sequence 76, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirst, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Gooddearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350NIM
; CURRENT APPLICATION NUMBER: US/09/759.130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-76

Query Match 12.8%; Score 28; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 4.5e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 GSHISGRGFLTYASSDHPDLITCLER 127
DB 100 GSHISGRGFLTYASSDHPDLITCLER 127

RESULT 5
US-09-759-130B-75
; Sequence 75, Application US/09759130B
; Publication No. US20030022279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarthy, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D

```

; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirtel, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodearl, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; FILE REFERENCE: MP100-5350ANIM
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-759-130B-75

Query Match      12.8%; Score 28; DB 9; Length 681;
Best Local Similarity 100.0%; Pred. No. 7e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 GSGHISGRGFLTYASSDHPDLITCLER 127
DB 100 GSGHISGRGFLTYASSDHPDLITCLER 127

RESULT 6
US-10-309-389-2
; Sequence 2, Application US/10309389
; Publication No. US20030087396A1
; GENERAL INFORMATION:
; APPLICANT: Saito, Masaki
; TITLE OF INVENTION: Sialyltransferase and DNA encoding the same
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/309,389
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: 09/425,488
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-184184
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: US 09/112,563
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: JP 11-148603
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 359
; TYPE: PRT

```

```

; ORGANISM: Mus musculus
; US-10-309-389-2

Query Match      3.2%; Score 7; DB 10; Length 359;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 VTVLFKS 100
DB 206 VTVLFKS 212

RESULT 7
US-09-753-143-77
; Sequence 77, Application US/09753143
; Patent No. US20020102550A1
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELIS, JAMES GERMAN, AND JOANNA
; GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/753,143
; FILING DATE: 02-Jan-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/175,828
; FILING DATE: 1998-10-20
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: <Unknown>
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 77:
; US-09-753-143-77

Query Match      3.2%; Score 7; DB 10; Length 370;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 LQSKGIS 192
DB 241 LQSKGIS 247

```

```
RESULT 8
US-09-712-363-148
; Sequence 148, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07/419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-148

Query Match
Best Local Similarity 3.2%; Score 7; DB 9; Length 385;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 LRLRGD 52
|||
Db 280 LRLRGD 286

RESULT 9
US-10-253-007-56
; Sequence 56, Application US/10253007
; Publication No. US20030088073A1
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-074-999
; CURRENT APPLICATION NUMBER: US/10/253,007
; PRIOR FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: US/09/186,188
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Plant
```

```
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(524)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-10-253-007-56

Query Match
Best Local Similarity 3.2%; Score 7; DB 9; Length 524;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 LRLGDLN 54
|||
Db 335 LRLGDLN 341

RESULT 10
US-10-045-170A-1
; Sequence 1, Application US/10045170A
; Patent No. US20020164809A1
; GENERAL INFORMATION:
; APPLICANT: Qiu, Yongchang
; APPLICANT: Wang, Jack
; APPLICANT: Hewick, Rodney
; TITLE OF INVENTION: ACID-LABILE ISOTOPE-CODED EXTRACTANT (ALICE) AND ITS USE IN GC
; FILE REFERENCE: G15412AUSA
; CURRENT APPLICATION NUMBER: US/10/045,170A
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242643
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Bovine Serum Albumin
US-10-045-170A-1

Query Match
Best Local Similarity 3.2%; Score 7; DB 9; Length 604;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 LIFSST 71
|||
Db 11 LIFSST 17

RESULT 11
US-09-815-242-10414
; Sequence 10414, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,127
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 10414
;; LENGTH: 610
;; TYPE: PRT
;; ORGANISM: Escherichia coli
US-09-815-242-10414

Query Match
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 LQSKGIS 192
Db 258 LQSKGIS 264
```

```
RESULT 12
US-09-864-761-47749
; Sequence 47749, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
```

```
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 47749
;; LENGTH: 100
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL109809.9
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.88
;; OTHER INFORMATION: EST HUMAN HIT: AM406955.1, EVALUE 1.00e-01
;; OTHER INFORMATION: SWISSPROT HIT: P01871, EVALUE 7.00e-03
US-09-864-761-47749
```

```
Query Match
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 LTVASS 116
Db 11 LTVASS 16
```

```
RESULT 13
US-09-911-777-5
; Sequence 5, Application US/09911777
; Patent No. US20020037852A1
; GENERAL INFORMATION:
; APPLICANT: BIOGEN, INC.
; APPLICANT: APOTECHE S.A.
; APPLICANT: BROWNING, Jeffrey
; APPLICANT: AMBROSE, Christine
; APPLICANT: MACKAY, Fabienne
; APPLICANT: TSCHOPE, Jurg
; APPLICANT: SCHNEIDER, Pascal
; TITLE OF INVENTION: BAFF, Inhibitors Thereof and Their Use
; FILE REFERENCE: A070 US
; CURRENT APPLICATION NUMBER: US/09/911,777
; CURRENT FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/117,169
; PRIOR FILING DATE: 1999-01-25
; PRIOR APPLICATION NUMBER: 60/143,228
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: PCT/US00/01788
; PRIOR FILING DATE: 2000-01-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-911-777-5

Query Match
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 LLANGV 202
Db 28 LLANGV 33

RESULT 14
US-09-864-761-47562
; Sequence 47562, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
```

FILE REFERENCE: Aecm1ca-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 47562
LENGTH: 107
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006023.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL - 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL - 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL - 1.6
OTHER INFORMATION: EST_HUMAN HIT: AW799294.1, EVALUATE 2.00e-47
OTHER INFORMATION: SWISSPROT HIT: P23301, EVALUATE 8.80e-01
US-09-864-761-47562

Query Match 2.7%, Score 6, DB 10, Length 107;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 184 NLSOK 189
Db 18 NLSOK 23

RESULT 15
US-09-910-150-21
Sequence 21, Application US/09910150
Patent No. US2002006898A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
APPLICANT: Rudolph-Owen, Laura
APPLICANT: Kapeller-Libermann, Rosana
APPLICANT: Tsai, Fong Ying

TITLE OF INVENTION: 13237, 18480, 2245 OR 16228 NOVEL HUMAN
FILE OF INVENTION: PROTEIN KINASE MOLECULES AND USES THEREFOR
FILE REFERENCE: 38155-20020.00
CURRENT APPLICATION NUMBER: US/09/910,150
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/219,028
PRIOR FILING DATE: 2000-07-18
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 108
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid
US-09-910-150-21

Query Match 2.7%, Score 6, DB 10, Length 108;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 198 LANGVL 203
Db 100 LANGVL 105

Search completed: May 15, 2003, 13:26:36
Job time : 12.6892 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 45.4269 Seconds
(without alignments)
1475.448 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGPSVTLALPVCAP.....YSAPRNGIAPLNOTAFALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*
- 9: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2659	99.6	503	21	AA19126
2	1924.5	72.1	715	22	AAU00670
3	1309.5	49.0	539	22	AAU00630
4	1309.5	49.0	586	22	AAU00629
5	1154.5	43.2	487	22	AAU00628
6	715	26.8	398	23	AAE22715
7	715	26.8	398	23	AAU79459
8	553.5	20.7	583	22	AAE75450
9	532	19.9	385	23	AAE22716
10	532	19.9	385	23	AAU79460

11	532	19.9	385	23	ABE97386
12	523	19.6	365	23	AAE22721
13	436.5	16.3	669	21	AAV70539
14	181.5	6.8	671	23	AAE14565
15	181.5	6.8	686	23	AAE14568
16	177.5	6.6	686	23	AAE14564
17	170.5	6.4	75	22	ABE42217
18	170.5	6.4	75	22	AAE63102
19	170.5	6.4	75	22	AAV75913
20	170.5	6.4	75	22	AAV36024
21	170.5	6.4	75	23	ABE45349
22	170.5	6.4	415	20	AAV13361
23	170.5	6.4	415	21	AAE24404
24	170.5	6.4	415	22	AAE88559
25	169	6.3	372	22	AAE80229
26	168	6.3	415	22	AAE88415
27	164.5	6.2	415	20	AAV31743
28	164.5	6.2	419	22	ABE22064
29	162.5	6.1	1733	23	ABE06023
30	162.5	6.1	3631	22	ABE28277
31	161.5	6.0	3623	20	AAV27020
32	159.5	6.0	110	20	AAV27042
33	157	5.9	315	22	AAE75323
34	154	5.8	464	22	AAE93631
35	153.5	5.7	728	22	AAE47559
36	153.5	5.7	689	22	AAE85060
37	150.5	5.6	690	22	AAU00713
38	150.5	5.6	717	22	AAU00196
39	150.5	5.6	718	22	AAU00200
40	150.5	5.6	1083	22	AAU00714
41	150.5	5.6	1084	22	AAU00718
42	150.5	5.6	1111	22	AAU00197
43	150.5	5.6	1112	22	AAU00201
44	150.5	5.6	1192	22	AAU00712
45	150.5	5.6	1192	22	AAU00712

ALIGNMENTS

RESULT 1	AA19126
ID	AA19126 standard; Protein; 503 AA.
XX	AA19126;
AC	AA19126;
XX	19-FEB-2001 (first entry)
DT	19-FEB-2001 (first entry)
XX	Polypeptide isolated from lymph node stromal cells of fsn -/- mice.
DE	Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW	immune system disorder; cancer; viral infection; HIV infection;
KW	blood vessel growth; tumour necrosis factor disorder; arthritis;
KW	inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW	cardiac failure.
XX	Mus sp.
XX	WO200058463-A1.
PN	05-OCT-2000.
PD	18-FEB-2000; 2000WO-NZ00015.
PF	25-MAR-1999; 99US-0276268.
PR	26-AUG-1999; 99US-0383586.
XX	(GENE-) GENESIS RES & DEV CORP LTD.
PA	Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI	Murison JG;
PI	WPI; 2000-664924/64.
XX	
DR	

DR N-PSDB: AAA96736.

XX polypeptide expressed in mammalian fan -/- lymph node stromal cells,
PT useful for modulating growth of blood cells, for treating inflammatory
PT and tumour necrosis factor-mediated disorders, cancer and viral
PT disorders

PS Claim 1; Page 68-69; 75pp; English.

XX The present sequence represents a polypeptide sequence which is
CC isolated from lymph node stromal cells of fan -/- mice. The
CC polynucleotides and their polypeptides are useful for treating an
CC inflammatory disorder, disorder of immune system and cancer selected
CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
CC viral disorder, in particular HIV infection and for modulating the
CC growth of blood vessels. The polypeptides are useful for treating a
CC tumour necrosis factor (TNF) mediated disorder, such as those selected
CC from arthritis, inflammatory bowel disease and cardiac failure and a
CC fibroblast growth factor-mediated disorder. It is also useful in assays
CC to determine biological activity, to raise antibodies, to isolate
CC corresponding ligands or receptors, to quantify levels of protein or
CC cognate corresponding ligand or receptors, as anti-inflammatory agents,
CC and in compositions for the treatment of skin, connective tissue and
CC immune system diseases. The polynucleotide is useful as marker for
CC tissue, as a chromosome marker or tags in the identification of a
CC genetic disorder.

XX Sequence 503 AA:

Query Match 99.6%; Score 2659; DB 21; Length 503;

Best Local Similarity 99.6%; Pred. No. 2.6e-243; Mismatches 1; Indels 0; Gaps 0;

Matches 501; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGTGAGGSPVATLALFVACAPRLQAEEELGDCGGHIVTSODSGTMTSKNPGTYPVYVCE 60
DB 1 MGTGAGGSPVATLALFVACAPRLQAEEELGDCGGHIVTSODSGTMTSKNPGTYPVYVCE 60
QY 61 KTTTPKPKRLRLRGDINTEKSTCASDYLLFSSATPDQYPCGSGWAPVKEKRLNSNVT 120
DB 61 KTTTPKPKRLRLRGDINTEKSTCASDYLLFSSATPDQYPCGSGWAPVKEKRLNSNVT 120
QY 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACGRODIAGDISGNT 180
DB 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACGRODIAGDISGNT 180
QY 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACGRODIAGDISGNT 180
DB 121 VLFKSGSHISGRFLTYAASDHPDLITCLERGSYFPEEKYSKPCPACGRODIAGDISGNT 180
QY 181 KDGYRDTSLCKAAIHAGITDELGHINLQSKGISHYEGGLANGVLSRHGSLSEKRL 240
DB 181 KDGYRDTSLCKAAIHAGITDELGHINLQSKGISHYEGGLANGVLSRHGSLSEKRL 240
QY 241 FTTPGMNTTVAIPSVIFAILLTGMCIFATCRKRRKGNPNVVSADQCTGCKWKQIKYF 300
DB 241 FTTPGMNTTVAIPSVIFAILLTGMCIFATCRKRRKGNPNVVSADQCTGCKWKQIKYF 300
QY 301 ARHSTFTTISYDNKEKMTOKLITSDMADYQOPLMIGTGVARRKSTFRMDTDEY 360
DB 301 ARHSTFTTISYDNKEKMTOKLITSDMADYQOPLMIGTGVARRKSTFRMDTDEY 360
QY 361 RYNTASGHYDCPRHPRGHEVALPLTHSPEYATPIVEHLLRAHTFTSGSGYVGGPP 420
DB 361 RYNTASGHYDCPRHPRGHEVALPLTHSPEYATPIVEHLLRAHTFTSGSGYVGGPP 420
QY 421 TRKHSHSGGPPATGANGVESYQRPAPKPYGGGYDPRASSFLDSRPAQSQMSTSG 480
DB 421 TRKHSHSGGPPATGANGVESYQRPAPKPYGGGYDPRASSFLDSRPAQSQMSTSG 480
QY 481 DDGYSAPRNGLAFLNQTAMTALL 503
DB 481 DDGYSAPRNGLAFLNQTAMTALL 503

RESULT 2
AAU00670
ID AAU00670 standard; Protein: 715 AA.

XX AAU00670;
XX 07-SEP-2001 (first entry)
DE Human TANGO 229 polypeptide.

XX Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;
KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW hypoglycaemia; diabetes mellitus; endometritis; pulmonary embolism;
KW muscular dystrophy; immuno-competence; vertebrate; blood; serum.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..34

FT /note= "Signal peptide"

FT Domain 35..455

FT /note= "Extracellular domain"

FT Protein 35..715

FT /note= "Mature human TANGO 229"

FT Domain 456..480

FT /note= "Transmembrane domain"

FT Domain 481..715

FT /note= "Cytoplasmic domain"

PN WO200129088-A1.

PD 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI: 2001-308477/32.

XX N-PSDB: AAS00660.

DR New isolated nucleic acid molecule for diagnosis, prevention, and

XX therapy of human and other animal disorder, or as modulating agent for

PT regulating cellular processes

XX Claim 8; Fig 1; 263pp; English.

PS The sequence represents human TANGO 229 polypeptide. This protein and

CC similar others exhibit the ability to affect growth, proliferation,

CC survival, differentiation, activity, morphology, or movement/migration

CC of, e.g., T cells and cells of the heart, liver, pancreas, placenta,

CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral

CC blood leukocyte, bone marrow or thymus tissue. They can be used as

CC modulating agents for regulating cellular processes, thus, the proteins

CC and their associated nucleic acids can be used to prognosticate, prevent,

CC diagnose, or treat disorders associated with physiological processes.

CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,

CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,

CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention

CC deficit disorder, Crohn's disease, gastroenteritis, goitre,

CC hypoglycaemia, diabetes mellitus, endometritis, pulmonary embolism and

CC muscular dystrophy. Antibodies to disorders such as these can be made by

CC providing a polypeptide of the invention to an immuno-competent

CC vertebrate and harvesting blood or serum from the vertebrate.

XX Sequence 715 AA:

Query Match 72.1%; Score 1924.5; DB 22; Length 715;
Best Local Similarity 55.8%; Pred. No. 2.4e-173;

CC Inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These
 CC host cells allow for the identification of compounds that bind to NHP
 CC receptors or trigger NHP-mediated pathways.

XX Sequence 487 AA:

Query Match 43.2%; Score 1154.5; DB 22; Length 487;
 Best Local Similarity 48.8%; Pred. No. 1.7e-100;
 Matches 237; Conservative 24; Mismatches 26; Indels 199; Gaps 1;

QY 44 MTSKNPTGTPNTVCEKITTPVKGRLILRLDNLNESTKASDYLLEFSATQTPYC 103
 DB 1 MTSKNPTGTPNTVCEKITTPVKGRLILRLDNLNESTKASDYLLEFSATQTPYC 60
 QY 104 GSNAPKELRLNSNEVTLFKSGSHIGRGFLTYASSDPDLITCLERSHYPEEYRSK 163
 DB 61 GSNAPKELRLNSNEVTLFKSGSHIGRGFLTYASSDPDLITCLERSHYPEEYRSK 120
 QY 164 FCPAGCRDIADISGNTRKDYRDTSLCKAAIHAGIITDELGCHINLQSGKISHYGLL 223
 DB 121 FCPAGCRDIADISGNTRKDYRDTSLCKAAIHAGIITDELGCHINLQSGKISHYGLL 180
 QY 224 ANGVLSHSGLSERKRLF----- 241
 DB 181 ANGVLSHSGLSERKRLF----- 240
 QY 242 ----- 241
 DB 241 QDGPSSWASDSSNNHPRMLEIDGKKKIGIRTTGSGSNFNYVSVANNEKNN 300
 QY 242 ----- 241
 DB 301 SKMKTYGIVNNEKVFQGSNFRDPVQNNFIPYAVYRVVPQWHORIALKEVLIGC 360
 QY 242 ----- 264
 DB 361 QITQGNDSLWKRKTSOSTSVSTKKEDETTTRIPSEETSTGINTTYAIPVLVAVFA 420
 QY 265 GMGIFALICRRKKRGPNVSADQKTCMKQIKYFARHOSTEFTISYDNEKEWTKRLD 324
 DB 421 GMGIFALICRRKKRGPNVSADQKTCMKQIKYFARHOSTEFTISYDNEKEWTKRLD 480
 QY 325 ITSDMA 330
 DB 481 ITSDMA 486

RESULT 6
 AAE22715
 ID AAE22715 standard; Protein; 398 AA.
 XX
 AC AAE22715;
 XX
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy1 protein.
 XX
 KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopoiesis; memory; platelet; plastic anaemia; antiinflammatory;
 KW opthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;

KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 KW immunosuppressive; chromosome 6q21.

OS Homo sapiens.

PN W0200222815-A1.

PD 21-MAR-2002.

PE 12-SEP-2001; 2001MO-US28488.

PR 11-SEP-2000; 2000US-0659671.

PR 06-SEP-2001; 2001US-317902P.

PA (HYSE-) HYSEQ INC.

PI Tang YT;

DR WPI: 2002-393966/42.

DR N-PSDB; AAD35992.

PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects

PS Claim 3; Page 123-125; 152pp; English.

CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy1
 CC protein. Neuropilin-Hy1 gene is located on chromosome 6q21.

SO Sequence 398 AA:

Query Match 26.8%; Score 715; DB 23; Length 398;
 Best Local Similarity 54.3%; Pred. No. 6.7e-59;
 Matches 159; Conservative 36; Mismatches 60; Indels 38; Gaps 8;

QY 29 GDGCGHIVTSODSGTMTSKNYPGTPYNTVCEKITTPVKGRLILRLDNLNESTKASD 88
 DB 4 GDGCGHIVTSODSGTMTSKNYPGTPYNTVCEKITTPVKGRLILRLDNLNESTKASD 63
 QY 89 YLFFSATDQYGPYCGSNAPKELRLNSNEVTLFKSGSHISGRGFLTYASSDPDLIT 148
 DB 64 YLFFSATDQYGPYCGSNAPKELRLNSNEVTLFKSGSHISGRGFLTYASSDPDLIT 121
 QY 149 CLENGSHYFEKYSK-FCPAGCRDIADISGNTRKDYRDTSLCKAAIHAGIITDELGCH 207

Db 122 QGDRPSEKTLDOQSRTFLATGTFVFKDSFS--TDG---TSLCKAAAHAGIADDELGQ 175
 QY 208 INLQSKGISHEGLANGVLSRHGSLSEKRR---LFTPGMNTTYVAIPSVIRALL 263
 Db 176 ISVLQKRGISREGILANGVLSREFEIRREQLFSSVLTYSMGNTHAV-----TELMF 228
 QY 264 TGMGIFAICRRK-----KKGNPYVSADAK-----TGCKMKQIKY 298
 Db 229 PHMIVHSGKTRREGSIAAEEGVPRLYVIOKQELVODLVAVATGCSRSLSF 281
 RESULT 7
 AAU79459
 ID AAU79459 standard; Protein; 398 AA.
 AC AAU79459;
 XX
 XX
 DT 15-JUL-2002 (first entry)
 XX
 XX Human Neuropilin-Hy1.
 DE
 XX Human: neuropilin-Hy1; chromosome 6q21; neuronal growth;
 KW nerve regeneration; neurodegenerative disease; learning disorder;
 KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
 KW organ growth; nervous system lesion; cancer; cell proliferation;
 KW cell differentiation; stem cell growth factor activity;
 KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
 KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
 KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
 KW reperfusion; food supplement; DNA microarray.
 XX
 XX Homo sapiens.
 OS
 XX WO200222780-A2.
 PN
 XX 21-MAR-2002.
 PD
 XX 11-SEP-2001; 2001WO-US28590.
 PF
 XX 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-0659671.
 XX
 XX (TANG/) TANG T Y.
 PA
 XX Tang TY;
 PI
 XX WPI: 2002-351881/38.
 DR N-PSDB; ABK49565.
 XX
 XX New neuropilin-like polypeptides for diagnosing, preventing and
 PT treating neurological conditions and disorders, cancers, and for
 PT inducing angiogenesis and neovascularisation
 XX
 PS Claim 3; Page 118-120; 144pp; English.
 CC The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC mapping genetic neuronal defects and degenerative diseases like
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischemic lesions, infectious lesions,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for

CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopaenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hy1.
 CC
 SQ Sequence 398 AA;
 Query Match 26.8%; Score 715; DB 23; Length 398;
 Best Local Similarity 54.3%; Pred. No. 6.7e-59;
 Matches 159; Conservative 36; Mismatches 60; Indels 38; Gaps 8;
 QY 29 GDGCHIVTSODSGTMTSKNTPGTYPNYTCCKITVYKGRLLIRLDGDNIESKTCASD 88
 Db 4 GDGCHLVITYDSDGTMSTKNTPGTYPNHYTCCKITVYKGRLLIRLDGDIESTQASD 63
 QY 89 YLLFSSATDQYQPYGCSMAVPKELNLNSNEVTLPKSGSHISGRGFLTTYASDHPDLIT 148
 Db 64 YLLFTSSDQYQPYGCSMTVPKELNLNTEVTLPKSGSHISGRGFLTTYASDHPD--S 121
 QY 149 CLERSHVFEEYK-RCPACGRDIAGDISGTRKGYRDTSLCKAAAHAGITDELGH 207
 Db 122 QGDRPSEKTLDOQSRTFLATGTFVFKDSFS--TDG---TSLCKAAAHAGIADDELGQ 175
 QY 208 INLQSKGISHEGLANGVLSRHGSLSEKRR---LFTPGMNTTYVAIPSVIRALL 263
 Db 176 ISVLQKRGISREGILANGVLSREFEIRREQLFSSVLTYSMGNTHAV-----TELMF 228
 QY 264 TGMGIFAICRRK-----KKGNPYVSADAK-----TGCKMKQIKY 298
 Db 229 PHMIVHSGKTRREGSIAAEEGVPRLYVIOKQELVODLVAVATGCSRSLSF 281
 RESULT 8
 AAG75450
 ID AAG75450 standard; Protein; 583 AA.
 AC AAG75450;
 XX
 XX 03-SEP-2001 (first entry)
 DT
 XX
 XX Human colon cancer antigen protein SEQ ID NO:6214.
 DE
 XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.
 KW
 XX Homo sapiens.
 OS
 XX WO200122920-A2.
 PN
 XX 05-APR-2001.
 PD
 XX 28-SEP-2000; 2000WO-US26524.
 PF
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 XX (HUMA-) HUMAN GENOME SCL INC.
 PA
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI WPI: 2001-23537/24.
 DR N-PSDB; AAH34855.
 XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11; Page 7657-7660; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 XX Sequence 583 AA;
 SQ
 Query Match 20.7%; Score 553.5; DB 22; Length 583;
 Best Local Similarity 30.3%; Pred. No. 2.6e-43;
 Matches 163; Conservative 82; Mismatches 194; Indels 99; Gaps 19;
 XX
 OY 8 PSYVALLFAVCAPRLRLAELGDCGHIYVSODSGTMSKNYPTGYPTVCEKIIYTPK 67
 DB 26 PLFLLTLVLLLEDDAGAGGDCGHTVGLPESGTLISINYPOTYNSFCWEIIRVNM 85
 OY 68 GKRLILRLGLNIE-SKTQASDYILFSS----ATDYGPICG-SWAPYKEIRLNSNEYTV 121
 DB 86 GERVRIRKFGFDIEDSDSCHFNRYIRYNGIGVSRTEIGKTCGLGLQNMHSIESGNIITL 145
 OY 122 LFKSGSHISGRGFLTYVASSDHPDLITCLERSHYFEFKSKFPCACRIADISGNTK 181
 DB 146 LFMGSHVSGRGLASTSVLDKODLITCLDPTASFLPEPSKYCPACCLLPFAEISGTP 205
 OY 182 DGYRDTSLCKAAIHAGITDELGHINLQSKGISHYEGLANGVLNRHGSLSERKFLP 241
 DB 206 HGRDSSPLCMAGVHAGVSNLTGGQISVYISKIPYESSLANNVTSVGHLSLSTFTF 265
 OY 242 TTP-----GMITTVAIPIVIFIAL--LIGM-GIPIACRRKKKGNPNYVSADAKT 290
 DB 266 KTSGCYGTLMEGVIADPQITASSVLEWTDHGOENSMKPKARLKRKPPMAAFATDE 325
 OY 291 GCKMOKIYPPARHSGTEFTSYDNEKEMTQKLDLITSDMDYQPLMIGTIV-----AR 345
 DB 326 YQWLQ-----IDLNKKKIT--GITTTGITWENNYVSAVRILYSDDGQ 368
 OY 346 KGSFPRMDTEFEVRNTEASGHYDCPHRPGRHEVALPLTHSEPEVATPIVERHLLRAH 405
 DB 369 KMVYVRPGEVQDKI-----FQGNKDY-HQDVNRKF-LP-----PIAR-FIRVN 410
 OY 406 TRESTQSYRY-----PCRPPIKSHSHSGGPPATGATQVESYQ 444
 DB 411 PTWQOQRIAKMELLGCOPIPKGRPPRLTPPPRNSNDKNTTA--PPRIAKRAKRF 468
 OY 445 RPASPKPVGGYDKPAAASFIDNSDPAOSOMTSGGDDGYSAPRNGLAGLNOTAMTL 502
 DB 469 QPIQPR-----SSNEF-----PAQTEQTASP-----IRNTVTYTNVTDVAL 507

XX
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy2 protein.
 XX
 KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnerability; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH Peptide 1..20
 FT /label= Signal_peptide
 FT Protein 61..385
 FT /note= "Human mature neuropilin-Hy2 protein"
 XX
 XX NO200222815-A1.
 XX
 XX 21-MAR-2002.
 XX
 XX 12-SEP-2001; 2001MO-US28488.
 XX
 XX 11-SEP-2000; 2000US-0659671.
 XX 06-SEP-2001; 2001US-317902P.
 XX
 XX (HYSE-) HYSDQ INC.
 XX
 XX Tang YT;
 XX WPI, 2002-393966/42.
 XX N-PDB; AAD35994.
 DR
 DR Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects -
 XX
 XX Claim 3; Page 128-130; 152pp; English.
 XX
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes

CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
CC protein. Neuropilin-Hy2 gene is located on chromosome 6q21.
XX
XX
SQ Sequence 385 AA;

Query Match 19.9%; Score 532; DB 23; Length 385;
Best Local Similarity 45.5%; Pred. No. 1.5e-41;
Matches 111; Conservative 38; Mismatches 89; Indels 6; Gaps 3;

OY 8 PSLALLFVAVCAPIRLQAEELDGGCHIVTSODSGMTSKNPGYPPNTVCEKIIITPK 67
DB 2 PFLULLLVLLLLLEDAGAQGGDGGCHIVLPESGTLISINPOTYPNSTVCEWEIRVM 61
OY 68 GKRLILRLGLDLINE-SKTCASDYLLFSS---ATDQGYPCG-SVAWPKELRLNENEVY 121
DB 62 GERVRIKRGDPDIEDSDSCHFNLYRLTYNGICVSRFEIKYCCGLGOMHNSISKNETTL 121
OY 122 LFKSGSHISGRGFLITYASSDHPDLITCLERGSHTFEERYSKFCPCAGCDIAGDISGNTK 181
DB 122 LFMSSIHVSGRGFLITYASVIDKQDITCLDTPASNFLPEFSKYCPAGCLLPAEISGITP 181
OY 182 DGYRPTSLCKRAIHAGIITDELGCHINLQSKGISHYEGLLANGVLSRHGSLSEKRLF 241
DB 182 HGVRDSSPLCMAGVHAGVAVSNLGGQISVVISKGIPIYESSLIANNVTSVGHLSLSTPTF 241
OY 242 TTPG 245
DB 242 KTSG 245

RESULT 10
AAU79460
ID AAU79460 standard; Protein: 385 AA.

AAU79460;

15-JUL-2002 (first entry)

Human Neuropilin-Hy2.

Human: neuropilin-Hy2; chromosome 6q21; neuronal growth;
neurodegeneration; neurodegenerative disease; learning disorder;
memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
organ growth; nervous system lesion; cancer; cell proliferation;
cell differentiation; stem cell growth factor activity;
Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
reperfusion; food supplement; DNA microarray.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..20

Protein /label= Signal_peptide
/label= Mature_neuropilin_Hy2

MO200222780-A2.

21-MAR-2002.

11-SEP-2001; 2001WO-US28590.

11-SEP-2000; 2000US-0659671.

06-SEP-2001; 2001US-0659671.

(TANG/) TANG T Y.
Tang TY;

DR WPI: 2002-351881/38.
DR N-PSDB; ABR49567.
XX
XX
XX New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation.

Claim 3; Page 123-125; 144pp; English.

The invention relates to an isolated neuropilin-like polypeptide
including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
encoding the proteins and the coding regions of the cDNAs. Also included
is a nucleic acid array comprising the cDNAs attached to a surface used
for detecting full-matches or mismatches to the cDNAs. The genes
for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
The nucleic acid array is useful for detecting full-matches or mismatches
to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
in modulating neuronal growth, regenerative capacity, treating
neurodegenerative diseases, learning and memory disorders, diagnosing and
mapping genetic neuronal defects and degenerative diseases like
Alzheimer's disease, for inducing angiogenesis, and neovascularisation
and organ growth and development (e.g. the heart). The nervous system
disorders include lesions of central or peripheral nervous systems,
including traumatic lesions, ischemic lesions, infectious lesions,
degenerative lesions, lesions associated with nutritional diseases or
disorders, neurological lesions, and lesions caused by toxic substances.
The neuropilin-like proteins and cDNAs are also useful as markers for
cancers. The neuropilin-like proteins are useful for regulating cell
proliferation, cell differentiation, stem cell growth factor activity,
for inducing proliferation of neural cells, regeneration of nerve and
brain tissue, for treatment of central and peripheral nervous system
diseases, and neuropathies, such as Parkinson's disease,
Huntington's disease, amyotrophic lateral sclerosis, to regulate
haematopoiesis and treat myeloid and lymphoid cell disorders, various
anaemias, and platelet disorders, such as thrombocytopaenia,
regeneration and treatment of lung or liver fibrosis, reperfusion
injury in various tissues and as a food supplement or molecular
weight marker. The cDNAs are useful in gene identification, genome
mapping, transgenics, as hybridisation probes, for primer design, for
gene chips and as a DNA antigen. The present sequence represents
neuropilin-Hy2.

Sequence 385 AA;

Query Match 19.9%; Score 532; DB 23; Length 385;

Best Local Similarity 45.5%; Pred. No. 1.5e-41;

Matches 111; Conservative 38; Mismatches 89; Indels 6; Gaps 3;

OY 8 PSLALLFVAVCAPIRLQAEELDGGCHIVTSODSGMTSKNPGYPPNTVCEKIIITPK 67
DB 2 PFLULLLVLLLLLEDAGAQGGDGGCHIVLPESGTLISINPOTYPNSTVCEWEIRVM 61
OY 68 GKRLILRLGLDLINE-SKTCASDYLLFSS---ATDQGYPCG-SVAWPKELRLNENEVY 121
DB 62 GERVRIKRGDPDIEDSDSCHFNLYRLTYNGICVSRFEIKYCCGLGOMHNSISKNETTL 121
OY 122 LFKSGSHISGRGFLITYASSDHPDLITCLERGSHTFEERYSKFCPCAGCDIAGDISGNTK 181
DB 122 LFMSSIHVSGRGFLITYASVIDKQDITCLDTPASNFLPEFSKYCPAGCLLPAEISGITP 181
OY 182 DGYRPTSLCKRAIHAGIITDELGCHINLQSKGISHYEGLLANGVLSRHGSLSEKRLF 241
DB 182 HGVRDSSPLCMAGVHAGVAVSNLGGQISVVISKGIPIYESSLIANNVTSVGHLSLSTPTF 241
OY 242 TTPG 245
DB 242 KTSG 245

RESULT 11
ABB97386
ID ABB97386 standard; Protein: 385 AA.

AC ABB97386;
 XX
 DT 27-JUN-2002 (first entry)
 XX
 DE Novel human protein SEQ ID NO: 654.
 XX
 KW Human; antihaemic; vulnery; antiinflammatory; immunomodulator;
 KW antileukemia; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
 KW neuroprotective; antiparkinsonian; protein therapy; EST;
 XX expressed sequence tag.
 XX
 OS Homo sapiens.
 PN WO200222660-A2.
 XX
 PD 21-MAR-2002.
 XX
 PF 10-SEP-2001; 2001MO-US26015.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 XX
 PA (HYSE-) HYSBO INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Weinman T, Drmanac RT;
 XX
 DR WPI: 2002-292408/33.
 DR N-PSDB; ABB32572.
 XX
 PT An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis -
 XX
 PS Claim 20; SEQ ID NO 654; 509bp; English.
 XX
 CC The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibit e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention.
 CC
 SQ Sequence 385 AA;
 Query Match 19.9%; Score 532; DB 23; Length 385;
 Best Local Similarity 43.5%; Pred. No. 1.5e-41;
 Matches 111; Conservative 38; Mismatches 89; Indels 6; Gaps 3;
 OY 8 PSYALIFAVCAFLRLQAEELAGGCGHYVTSODSGTMSKNTPGTYPNVYCKEITVPK 67
 DB 2 PLFLILLVLLLEEDAGAGGCGCHYVAGPSSGTLTINTPTPNSVCEWEIRVKM 61
 OY 68 GKRLIRLADLNTIE-SKTCASDYLFSS---ATDQYGPYC-SMAVPEKRLNSNEVY 121
 DB 62 GERVARKFGDFEDSDCFNYLRIYNGISVGRTEIGKCGLGLOMNHISEKNEITL 121
 OY 122 LFFSGSHIGRGFLTYASSDHPDLITCLERGHYFEERKSKPCPACGRDIADISGNTK 181
 DB 122 LFFSGSHIGRGFLTYASSDHPDLITCLERGHYFEERKSKPCPACGRDIADISGNTK 181
 OY 182 DGYRDTSLCKAIIHAGITDELGCHINILLOSCKISHYEGILANGVLSRHGSLSEKRF 241
 DB 182 HGYRDSPLCMGVHAGVSNITLGGQISVISCIPRYESSLANNTSVYGHLSITLFT 241
 OY 242 TTPG 245
 DB 242 KTSQ 245
 RESULT 12

AAE22721
 ID AAE22721 standard; Protein: 365 AA.
 AC AAE22721;
 XX
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy2 mature protein.
 XX
 KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virocid; antibacterial; cytoskeletal;
 KW immunosuppressive.
 XX
 OS Homo sapiens.
 PN WO200222815-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 12-SEP-2001; 2001MO-US28488.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-317902P.
 XX
 PA (HYSE-) HYSBO INC.
 XX
 PI Tang YT;
 XX
 DR WPI: 2002-393966/42.
 XX
 PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects -
 XX
 PS Disclosure; Page 131-132; 152pp; English.
 XX
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2

[illegible]

17-MAY-2002 (first entry)

Human MASP-2 protein, alternative version.

Human; mannan-binding lectin associated serine protease-2; MASP-2; MBL; complement fixation; infection; microbe; retrovirus; HIV; abortion; human immunodeficiency virus; pathogenic bacteria; inflammatory disorder.

Homo sapiens.

Key Location/Qualifiers
Peptide 1..15
Protein /label= Signal_peptide
16..686
/label= Mature_MASP-2_protein
Misc-difference 155
/note= "Encoded by CA"
Misc-difference 156
/note= "Encoded by C"
Misc-difference 298
/note= "Encoded by CAG"
Misc-difference 299
/note= "Encoded by CCT"

MO200206460-A2.

24-JAN-2002.

13-JUL-2001; 2001MO-DR00499.

13-JUL-2000; 2000DK-0001089.
01-JUN-2001; 2001DK-0000870.

(JENS/) JENSENIUS J C.
(THIE/) THIEL S.

Jensenius JC, Thiel S;
WPI: 2002-179791/23.
N-PSDB; AAD24224.

Use of a polypeptide comprising amino acid sequence derived from mannan-binding lectin associated serine protease-2 (MASP-2) for producing pharmaceutical composition, to treat bacterial, fungal, viral infections -

Claim 41; Page 71-73; 76pp; English.

The invention relates to use of a polypeptide derived from mannan-binding lectin (MBL) associated serine protease-2 (MASP-2) for producing a pharmaceutical composition. MASP-2 is a complement-fixing enzyme and involved in lectin pathway of complement activation. The pharmaceutical composition comprising MASP-2 is useful for treating infections caused by microbes such as fungus, yeast, retrovirus (e.g. human immunodeficiency virus (HIV)), or pathogenic bacteria which are resistant to at least one antibiotic medicament or multiresistant. The polynucleotide encoding MASP-2 is useful for treating patients deficient in MASP-2. The invention also discloses MASP-2 assays which are useful for determination of MASP-2 activity or levels in patients suffering from e.g. infections, inflammatory disorders and spontaneous recurrent abortion. The pharmaceutical composition comprising MASP-2 inhibitor is useful for treating inflammatory disorders. The present sequence is human MASP-2 protein.

Note: The present sequence is stated as being the same as SEQ ID NO:2 shown in figure 6 of the specification (AAE15564).

However the sequences differ at various locations.

Sequence 686 AA:

Query Match 6.88; Score 181.5; DB 23; Length 686;
Best Local Similarity 24.48; Pred. No. 7.1e-08;
Matches 52; Conservative 42; Mismatches 94; Indels 25; Gaps 7;

```

OY 32 CGHITSQDSGTWTSKNTPGTYPNVTCEKIITVPKRLIL---RLGDLNIESKT-CAS 87
Db 184 CSGQVFTORSSELSPPEYPRPKLSCTYSISLEGEFSVILDVESEFVETHPETLCBY 243
OY 88 DYLFSSATDOYGPYCGSMVAPKELRLNSNEVTLP---KSGSHISGCELTVAASDH- 143
Db 244 DFLKIQIDREHGPFGK-TLPHRIETKSNVTYITFVTDESGDH--TGWKIHYTSTAH 299
OY 144 -----PDLITCLERGRSHYEEKYSKFCPACGRDIAGDISGNT-----KGYRDTSL 190
Db 300 CPYMAPPENGHSPYQAKYILKDSFISFCETGYELLQGHLPKSFYAVOCQKDSWDRPMP 359
OY 191 CKAITHAGIITDELGCHINLQSKGISHEGLL 223
Db 360 ACSIVDCGPPDDLPSGRVEYITGPGVTYKAVI 392

```

Search completed: May 15, 2003, 13:20:46
 Job time : 49.4269 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 14.454 Seconds

(Without alignments)
3345.478 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671

Sequence: 1 MGTGAGGSPVALLFAVCAP.....YSAPRNGLAFLNQTAMTALL 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pir73: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	181.5	6.8	686	1 A59271	Ra-reactive factor
2	169.5	6.3	694	2 JG6554	complement subcomp
3	168	6.3	1524	2 T30337	polyprotein - Aflr1
4	167.5	6.3	3623	2 T08618	intrinsic factor-B
5	162.5	6.1	3623	2 T09456	intrinsic factor-B
6	158.5	5.9	695	1 JQ0508	complement subcomp
7	157	5.9	927	1 JQ0948	AS antigen precurs
8	151.5	5.7	699	1 I54763	Ra-reactive factor
9	151	5.7	688	1 C1HRS	complement subcomp
10	150	5.6	1594	2 T30549	heparin - radblt
11	147	5.5	705	1 C1HURB	complement subcomp
12	142.5	5.3	966	1 B58788	procollagen C-endo
13	142.5	5.3	991	2 I49540	procollagen C-endo
14	142	5.3	579	2 JG7629	membrane-type friz
15	142	5.3	2083	2 T42721	CRP-ductin-alpha p
16	141.5	5.3	1070	2 T31069	tollid-BMP-1 like
17	141	5.3	319	2 I51569	UVS-2 protein - Af
18	137	5.1	449	2 A55362	procollagen I C-pr
19	130.5	4.9	1290	2 A55190	ebnerin precursor
20	130.5	4.9	3871	2 T22812	hypothetical prote
21	127.5	4.8	707	2 JG2318	procollagen C-endo
22	127	4.8	855	2 JG7731	membrane-bound arg
23	127	4.8	855	2 JG7731	membrane-bound arg
24	126	4.7	402	2 JH0403	procollagen I C-pr
25	125.5	4.7	2403	2 A59386	sanko - human
26	125	4.7	1019	2 A38738	coagulation factor
27	124	4.6	1057	1 A59288	dorsal-ventral pat
28	123.5	4.6	823	1 A58788	procollagen C-endo
29	122	4.6	730	1 BMH01	procollagen C-endo

30	121.5	4.5	597	2 S71352	metalloproteinase
31	121.5	4.5	1004	2 T30338	oviductin (EC 3.4.
32	119	4.5	767	2 T30018	hypothetical prote
33	117.5	4.4	347	2 T20618	hypothetical prote
34	116	4.3	275	2 JG6506	tumor necrosis fac
35	115	4.3	277	2 A41735	hyaluronate-bindin
36	114	4.3	288	2 T33224	hypothetical prote
37	113.5	4.2	1464	2 A40136	fibropellin Ia - s
38	112.5	4.2	1846	2 S58984	development protei
39	111	4.2	846	2 H70599	hypothetical prote
40	111	4.2	1087	2 T31100	probable potassium
41	110.5	4.1	436	2 T23345	hypothetical prote
42	110	4.1	276	2 A47290	TSG-6 homolog PS4
43	110	4.1	504	2 S56745	mucin (clone PGM31
44	109.5	4.1	321	2 T33161	hypothetical prote
45	109.5	4.1	1571	2 T00062	hypothetical prote

ALIGNMENTS

RESULT 1

A59271
Ra-reactive factor (EC 3.4.21.-) 2 precursor - human
N:Alternate names: mannos binding protein-associated serine proteinase 2 (MASP-2)
C:Species: Homo sapiens (man)
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
C:Accession: A59271
R:Phlel, S.; Vorup-Jensen, T.; Stover, C.M.; Schwaebel, W.J.; Laursen, S.B.; Poulsen, Nature 386, 506-510, 1997
A:Title: A second serine protease associated with mannan-binding lectin that activate
A:Reference number: A59271; MUID:97242412; PMID:9087411
A:Accession: A59271
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-686 <JEN>
A:Cross-references: GB:Y09926; NID:94007626; PIDN:CAA71059.1; PID:94007627
A:Experimental source: tissue liver
A:Note: submitted to GenBank, December 1996
A:Note: parts of this sequence, including the amino end of the mature protein, were d
C:Genetics:
A:Gene: GDB:MASP2
A:Cross-references: GDB:6071500
A:Map position: 1p36.2-1p36.3
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement facto
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; hydrolase; serin
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-444,445-686/Product: Ra-reactive factor 2 #status predicted <MAT>
F:142-180/Domain: C1r/C1s repeat homology <EGF>
F:184-293/Domain: C1r/C1s repeat homology <C1R2>
F:300-361/Domain: complement factor H repeat homology <FH1>
F:366-430/Domain: complement factor H repeat homology <FH2>
F:445-679/Domain: trypsin homology <TRY>
F:72-90,142-156,152-165,167-180,184-211,241-259,300-348,328-361,366-412,396-430,434-5
F:158/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:444-445/Cleavage site: Arg-Ile (autolytic) #status predicted
F:443,532,633/Active site: His, Asp, Ser #status predicted

Query Match 6.88; Score 181.5; DB 1; Length 686;
Best Local Similarity 24.44; Pred. No. 7.3e-06;
Matches 52; Conservative 42; Mismatches 94; Indels 25; Gaps 7;

QY 32 CGHIVTSGDGTMTSKNVPGRVYTCERKITTPKGRRLT---RIGDNIESTK-CAS 87
DB 184 CCGGVFTORSGELSPPEPRPYRLSSCTYSISLEEGSVLLDFVESFDVETHPETLCPPY 243
QY 88 DYLLFSSATDQYGPYCGSMWPKELRLNSNEVYLF--KSGSHISRGFLFYASSDH- 143
DB 244 DFLNIQDRBHEHGFCK-TLPHRIETKSNVTITFTVDESGDH---TGMKIHITSTAH 299
QY 144 -----PDLITCLERSHFFEEKYKFCPCAGCDINDIGISGNT-----KDGRTDTSL 190

DB 300 CYPAPAPNGHYSPVQAKIILKSDPSIFCETGELLQGLPLKSTFAPVQKQKSDMPMP 359
QY 191 CKAIAHAGITDELGHINLQSKGISHYEGIL 223
DB 360 ACSIVDCGPPDDLPGSGRVETITGPGVTKYKAVI 392

RESULT 2
JC6554
Complement subcomponent C ISHAR GIF (EC 3.4.21.42) precursor [similarity] - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 03-Jun-2002
C:Accession: JC6554
R:Sakai, H.; Nakashima, S.; Yoshimura, S.; Nishimura, Y.; Sakai, N.; Nozawa, Y.
Gene 209, 87-94, 1998
A:Title: Molecular cloning of a cDNA encoding a serine protease homologous to complement
A:Reference number: JC6554; MUID:98192519; PMID:9524231
A:Accession: JC6554
A:Molecule type: mRNA
A:Residues: 1-694 <SAK>
A:Cross-references: DDBJ:D88250; NID:93080541; PIDN:BAA25797.1; PID:93080542
C:Comment: This protein is involved in glial cell differentiation and cartilage remodeling
C:Genetics:
A:Gene: r-98P
C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
C:Keywords: differentiation; glycoprotein; hydrolase; serine proteinase
F:1-31/Domain: signal sequence #status predicted <SIG>
F:117-133/Domain: C1r/C1s repeat homology <C1R>
F:12-694/Product: serine protease homolog #status predicted <MAT>
F:141-177/Domain: EGF homology <EGF>
F:300-360/Domain: complement factor H repeat homology <FHR>
F:444-681/Domain: trypsin homology <TRY>
F:180-412/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:481,535,637/Active site: His, Asp, Ser #status predicted

Query Match 6.3%; Score 169.5; DB 2; Length 694;
Best Local Similarity 28.0%; Pred. No. 6.3e-05;
Matches 52; Conservative 31; Mismatches 78; Indels 25; Gaps 7;

QY 15 FAVCAP---LRQAERLDGCGHIYTSQDSGMTSKNTGTPNTVCCKITTPKGRKL 71
DB 161 FCSCEPEYFLHDMRTCGNCSGDVFTALIGIASPNPNPENSRCYQIRLQEGFRL 220
QY 72 IL-FLGDLNIE---SKTCASDYLFFSATDQGYCGS-WAYPKRELNSNEVYLFKS 125
DB 221 VLTTRDFDVPADSEGCNDSLTPAAKNOQFGPCGFGPGLTKQSTLIDVFEPT 280
QY 126 GSHISRGFLLTYASSDHPDLITC-----LENGSHYEEKYKFTCPAGCRDING 174
DB 281 DLTGONKMGKRLRY---HDDPIPCPEKISANSIWEPEKAKYFKDVVKITVDGEFVVG 336
QY 175 DISGNT 180
DB 337 NV-GST 341

RESULT 3
730337
POLYPROTEIN - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: 730337
R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
Submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from X
A:Reference number: 730337
A:Accession: 730337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: EMBL:U81290; NID:92981640; PID:92981641; PIDN:AAC24717.1

Query Match 6.3%; Score 168; DB 2; Length 1524;

Best Local Similarity 30.0%; Pred. No. 0.00025;
Matches 45; Conservative 23; Mismatches 68; Indels 14; Gaps 5;

QY 24 QAEELGDCGCGHIVTSQDSGMTSKNYPGTYVNTVCCKITTPKGRKLRLRLGDLNIEBK 83
DB 972 QGPRYGSCEYILTSSEGVIESPNYLGNYPPDLHCQMRILNPAVKYLRVDLEKTEKD 1031
QY 84 T--CASDYLIES---SADDQGYPCGSMAVPEKELNSNEVYLFKSGSHISGRFLLTY 138
DB 1032 VSGCDDPLLVYNGIESKDLGVCGE--IRLSLSESESEITLPTNSNEVSGKGFSLKY 1090
QY 139 ASSDHPDLITCLERGSHYEEKSKF-CPA 167.
DB 1091 SPWDK-----QAQSKOLEENAAVGCBA 1113

RESULT 4
708618
Intrinsic factor-B12 receptor CUBILIN precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Aug-2002
C:Accession: T08618
R:Moestrup, S.K.; Kozyrakl, R.; Kristiansen, M.; Kaysen, J.H.; Rasmussen, H.H.; Brall
J. Biol. Chem. 273, 5235-5242, 1998
A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibody
A:Reference number: Z16459; MUID:98148073; PMID:9478979
A:Accession: T08618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3623 <MOE>
A:Cross-references: EMBL:AF022247; NID:93834379; PIDN:AAC71661.1; PID:93834380
C:Genetics:
A:Gene: CUBILIN
C:Superfamily: intrinsic factor-B12 receptor cubilin; EGF homology
C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-3623/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <MAT>
F:133-164/Domain: EGF homology <EGF1>
F:436-467/Domain: EGF homology <EGF>

Query Match 6.3%; Score 167.5; DB 2; Length 3623;
Best Local Similarity 20.7%; Pred. No. 0.00094;
Matches 110; Conservative 64; Mismatches 189; Indels 169; Gaps 26;

QY 32 CCHITTSQDSGMTSKNTGTPNTVCCKITTPKGRKLRLRLGDLNIE-SKTCASDYL 90
DB 932 CCEVLLTA-STGILIESPGRHNYPRGVNCTHWVVOGQIRLEFFSFLLEFHYNCTNDYL 990
QY 91 -LEFSATDQY-GPYGSMAVPEKELNSNEVYLFKSGSHISGRFLLTYASSD----- 142
DB 991 EYIDTAQOTFLGRYCGK-SIPPSLSNSNSIKLITVSDSALAHGCFSTNYEADSSVCL 1049
QY 143 -----HPDLITCLER-----GSHYEEKYKFTCPAGCRDIAGD 175
DB 1050 YDTDNFGMLSSPNPNPNYPSMWECIYRTVGLNQIALHFDPLFEDYFGQCYDVF-- 1107
QY 176 ISGNTRKGYRDTSLCKAIAIHAG-IITDELGHIN-----LQSKGISHY----- 219
DB 1108 ---EIRDGYETSPL--VGICYGSVLPPTIISHSKMLKFKSDAALAKGFSAYWDCSS 1162
QY 220 EGLLAN-----GVL-----SRHGLSKRR----- 239
DB 1163 TCGGGLNLTTPGVLSLPNTPMPIYHSSSECYWRLESHSGSPFELERQDFHLEHHPSCSLDY 1222
QY 240 ---LEFTPGMN-----ITVAIPSVIFIAL-----LTGNGIFATCRKRRKGN 280
DB 1223 LAVFEGPTTNSLIDKLGDTTPAPIRSNKQVYLLAKRTDAGQGRGEINRQACDN-- 1280
QY 281 PYVSADQKTCGQWKQIKYPPARHQSSTFTISYDNKEKMTQKIDLTSDMADYQOPLMTGT 340
DB 1281 --VIVNKTSGILIESINP-----NPYDKNQRCNMTIQAOTGTGMVNV----- 1320
QY 341 GTVAARKGSTRMDTDEEVNRTASGHIYDCPHRGRHEVALPLTHSEPEYATPIVERH 400

Db 1321 -----FFLGFDVES-YMNCSTVDYELDQGMKGRY-----CGNMPPGATYTSQHL 1367
 QY 401 LIRAHM---FSTOSGYRVPDPRTHKSHSSGGEPFPAWGATQYVESYORPASP 449
 Db 1368 VL-FHFDGINSKGRFM-----QWFTHCQGEMSGTAG-----SFSSPPIF 1408

RESULT 5
 T09456
 Intrinsic factor-B12 receptor Cubillin precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Aug-2002
 C:Accession: T09456
 R:Kozyrak, R.; Kristiansen, M.; Sliatkaroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N.
 Blood 91, 3593-3600, 1998
 A>Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characterization.
 A:Reference number: Z16677; MVID:98241400; PMID:9572993
 A:Accession: T09456
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3623 <KOZ>
 A:Cross-references: EMBL:AF034611; NID:93929528; PIDN:AAC82612.1; PID:93929529
 C:Genetics:
 A:Map position: 10p12
 C:Superfamily: Intrinsic factor-B12 receptor cubillin; EGF homology
 C:Keywords: receptor; vitamin B12 uptake
 F:1-24/Domain: signal sequence #status predicted <Sig>
 F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <MAT>
 F:436-467/Domain: EGF homology <EGF>

Query Match
 Best Local Similarity 6.1%; Score 162.5; DB 2; Length 3623;
 Matches 47; Conservative 24; Mismatches 57; Indels 37; Gaps 7;

QY 5 AGPSTALIFAVCAPLR-----LQAEELG-----DCGHITVS 38
 Db 2640 AACP---LMMRLCGPSCPLPLVPIPSQWIFVTVNEREHLGFLNAKYSFTDCGIG 2695

QY 39 ODSGTSTKNPGTYPNMYCEKIITVPKGRILRLRGDNIET-SKTCASDYLFF---S 93
 Db 2696 -DSGVITSPFNAYDSLTHCSKWLAPQGHITTLTFSDDIPHTTCAMDVTVNRNGS 2754

QY 94 SATDQYPCYGSWAVPKRELRLNSNEVTLFKSGSHISGRFLTY 138
 Db 2755 PESPITIGYCGN-SNPRITQSGSNQLVTFNSDHSILGGGFFATW 2798

RESULT 6
 S05008
 Complement subcomponent C 1SBAK.GIF (EC 3.4.21.42) precursor [similarity] - golden hamst
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
 C:Accession: S05008
 R:Kinoshita, H.; Sakiyama, H.; Tokunaga, K.; Imajobi-Ohts, S.; Hamada, Y.; Isono, K.; Sak
 FEBS Lett. 250, 411-415, 1989
 A>Title: Complete primary structure of a calcium-dependent serine proteinase capable of
 A:Reference number: S05008; MVID:89325606; PMID:2753140
 A:Accession: S05008
 A:Molecule type: mRNA
 A:Residues: 1-695 <KIN>
 A:Cross-references: EMBL:X16160; NID:949621; PIDN:CAA34286.1; PID:949622
 A>Note: part of this sequence, including the amino ends of both the heavy and light chain
 C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H
 C:Keywords: beta-hydroxyasparagine; calcium binding; duplication; glycoprotein; hydrolyas
 F:1-21/Domain: signal sequence #status predicted <Sig>
 F:17-133/Domain: C1r/C1s repeat homology <C1R1>
 F:122-444/Product: serine proteinase heavy chain #status experimental <HCH>
 F:141-177/Domain: EGF homology <EGF>
 F:181-293/Domain: C1r/C1s repeat homology <C1R2>
 F:300-360/Domain: complement factor H repeat homology <H1>
 F:365-428/Domain: complement factor H repeat homology <H2>

F:445-682/Domain: trypsin homology <TRY>
 F:446-695/Product: serine proteinase light chain #status experimental <LCH>
 F:71-89,141-153,149-162,164-177,181-208,240-257,300-347,327-360,365-410,392-428,432-5
 F:155/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:180,413/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:482,536,638/Active site: His, Asp, Ser #status predicted

Query Match
 Best Local Similarity 5.9%; Score 158.5; DB 1; Length 695;
 Matches 48; Conservative 37; Mismatches 76; Indels 25; Gaps 7;

QY 15 FAVCAP---LRLQAEELGDCGHITVSODSGTMTSKNYPGTYPNMYCEKIITVPKGRIL 71
 Db 161 FCSCPPEYFHDHMRKNGVNCSENVETALIGETISPPYPPYDENSRCFQILLEGFQV 220

QY 72 ILRL--GDNLIE---SKTCASDYLFPSSANDQYPCYGS-WAVPKRELRLNSNEVTLFKS 125
 Db 221 VTIQREDFVEPADSGNCGQSDSLFLPAKNRQFPCGNGFPGPLTLETSHSLDITVFT 280

QY 126 GSHISGRFLTYASDHPDLITC-----LEKSHYFEKYSKPCAGCRDIAG 174
 Db 281 DLTEQKKGMKLR-----HGDPICPKREITANSWVAPERKATYVRKDVYKISCVDFEAVEG 336

QY 175 DISGNT 180
 Db 337 NV-GST 341

RESULT 7
 J00948
 A5 antigen precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: J00948
 R:Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
 Neuron 7, 295-307, 1991
 A>Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homol
 A:Reference number: J00466; MVID:91337458; PMID:1908252
 A:Accession: J00466
 A:Molecule type: mRNA
 A:Residues: 1-927 <TAK>
 A:Cross-references: GB:D10467; GB:D01077; NID:9222962; PIDN:BA01360.1; PID:9222963
 A:Experimental source: tadpole, brain
 A>Note: this protein has motifs homologous to complement components C1r and C1s and t
 C:Comment: This protein is a neuronal cell surface molecule involved in the neuronal
 C:Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-termi
 C:Keywords: duplication; glycoprotein; transmembrane protein
 F:1-21/Domain: signal sequence #status predicted <Sig>
 F:22-927/Product: A5 antigen #status predicted <A5A>
 F:27-138/Domain: C1r/C1s repeat homology <C1R1>
 F:147-262/Domain: C1r/C1s repeat homology <C1R2>
 F:274-424/Domain: discoidin I amino-terminal homology <DN1>
 F:446-812/Domain: discoidin I amino-terminal homology <DN2>
 F:861-883/Domain: transmembrane #status predicted <TM>
 F:150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match
 Best Local Similarity 5.9%; Score 157; DB 1; Length 927;
 Matches 99; Conservative 80; Mismatches 177; Indels 152; Gaps 25;

QY 21 LRLQAEELGDCGHITVSODSGTMTSKNYPGTYPNMYCEKIITVPKGRILRLDLNT 80
 Db 136 IREYVKTGREGCRNFTS-SNGVIRKPKYDEKXPNALCEYIILFAKMGQIVLEFSPFL 194

QY 81 ES-----KTCASDYL---LFSATDQYPCYGSWAVPKRELRLNSNEVTLFKSGSHI 129
 Db 195 EADSNAPEGQTCRYDWLGIMDGFPVGPPIHGRYCGO-NTPGRVRSFTGLISMIFHTDSAI 253

QY 130 SGSGFLTYASSDHPDLITCLEKSHYFEKYSKPCPAGCRDIAGISGNTKGYVDTS 189
 Db 254 AKGCFRANFSV-----VQSTEDDFQ----- 274

```

0Y 190 LCKAIIHGIITDELGGHINLLOKSGISHHEGLLANVLSRHSLSEKRELF-----TRQ 245
Db 275 -CKREL--GMSGEL-----HHDQISVSSQGYMNNASERLNYENGWTRQ 318
0Y 246 MNTTVAIPSVIFALLITGCGT-FAICRRKKK-----GNPYVS-ADAQR- 289
Db 319 EDYKWEHVOVDLEMLRVSQIGTGAISKEKTKRYKYSKYVDISSNGEIMVILKQGNKH 378
0Y 290 -----TGCWKQIKYPPFAHSHOSTEFT-----ISYDNEKEKMTQIDLITSDADYQOPLMIG 339
Db 379 LVFNGNDATDVYVRFPSPKPYITHFVRLRPVTWEN--GISLRFELYGCKITDOPCSMIG 436
0Y 340 --TGVVARKGT-TRPMDTD--TEEVRYNTEASGHYCCPHRPGHREALPLTISEP--- 391
Db 437 MVSGLISDSQITLASSOVDNRNVPRLARLYTSSRS-----WALPSSNTHPTYKE 484
0Y 392 -YATPIVYRHLIRAHTEFSTQSGYVPPRPRLPHKS-----HSSGFGFPATGATGV 440
Db 485 WLQDLDAEBEKLYR-----GVIIQGGK--KRENYKFMKFKLGYENNG-----TEW 527
0Y 441 ESYQRPASPKPV-----GGGYKPKPAASSF 464
Db 528 EMIIMDSKNNKRTFEGCNTNTYTPELRTP 555

```

RESULT 8

Ra-reactive factor (EC 3.4.21.-) 1 precursor - human
 N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)
 C:Species: Homo sapiens (man)
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
 C:Accession: J14763; J08883
 R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.
 Int. Immunol. 6, 665-669, 1994
 A:Title: Molecular characterization of a novel serine protease involved in activation of
 A:Reference number: J54763; MUID:94289349; PMID:8018603
 A:Accession: J54763
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-699 <SAT>
 A:Cross-references: GB:D8593; NID:9790963; PIDN:BA005928.1; PID:g471128
 R:Ikada, F.; Takayama, Y.; Hatanue, H.; Kawakami, M.
 Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993
 A:Title: A new member of the C1s family of complement proteins found in a bactericidal
 A:Reference number: J08883; MUID:94059062; PMID:8240317
 A:Accession: J08883
 A:Molecule type: mRNA
 A:Residues: 1-234, 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>
 A:Cross-references: DDBJ:D15725; NID:9439712; PIDN:BA04477.1; PID:g439713
 A:Experimental source: liver
 C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 compo
 C:Genetics:
 A:Gene: GDB:MASP1; GDB:CRAP1; CRAP1; PRSS5; MASP
 A:Cross-references: GDB:361104; GDB:330954; OMIM:600521
 A:Map position: 3q27-3q28
 C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor F
 C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hydrolytic
 F:1-17/Domain: signal sequence #status predicted <SIG>
 F:18-448,449-699/Product: Ra-reactive factor #status predicted <MAT>
 F:19-135/Domain: C1r/C1s repeat homology <C1R1>
 F:143-181/Domain: EGF homology <EGF>
 F:185-294/Domain: C1r/C1s repeat homology <C1R2>
 F:301-362/Domain: complement factor H repeat homology <FHN>
 F:367-432/Domain: complement factor H repeat homology <FHN2>
 F:449-699/Domain: trypsin homology <TRY>
 F:449, 178, 407/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:73-91, 143-157, 153-166, 168-181, 185-212, 242-260, 301-349, 329-362, 367-414, 397-432, 436-572
 F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:448-449/Cleavage site: Arg-Tyr (autolytic) #status predicted
 F:490,552,646/Active site: His, Asp, Ser #status predicted

	Matches	48;	Conservative	45;	Mismatches	89;	Indels	27;	Gaps	8;
OY	32	CGHIVTSDDSGTMSKNTPGYTPNATVEKEIITVKGKRLILRLDL-NIESK---	TCAS	87						
Db	185	CSDNLEFRTGTYTSPDPPNPKNSECLYLTLEEGFMVNIQFEDIDIDIDHPPEPPY	244							
OY	88	DYLLSSATDQYGPYCGSMAVPEKRLNLSNEVYLFKSGSHISGRFLITY--ASSDAPD	145							
Db	245	DYIKIKVGPYKVLPGPCGEKA-PEPISYOSHVLILFHDSNAENNGMWLSRAAGNECEP	303							
OY	146	LITCL-----ERGSHYEERYKSKFCPAGCRDIADIGNT-----KGYRDIYSL-LCK	192							
Db	304	LQPPVHGKIEPSQAAKYFFKDDVLYVSCDYGKYVLLKDNVEDMTFOELCLDGTWSNKLPTCK	363							
OY	193	-----AAIHAGITDELGHINILNLSK	214							
Db	364	IYDCRPAPELIEHGLITFSTRNLLTYKSE	392							

RESULT 9

complement subcomponent C ISBAR.GIF (EC 3.4.21.42) precursor [validated] - human

N.alternate names: C1 esterase precursor

C.species: Homo sapiens (man)

C.date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Jun-2002

C.accession: A40496; A27381; S00224; S26732; S05634; A05140; A23396; A38407; B37820

R.kusumoto, H.; Hiroseawa, S.; Sailer, J.P.; Hagen, F.S.; Kurechi, K.

Proc. Natl. Acad. Sci. U.S.A. 85, 7307-7311, 1988

A.title: Human genes for complement components C1r and C1s in a close tail-to-tail ar

A.reference number: A40496; MUID:89017187; PMID:2459702

A.accession: A40496

A:molecule type: mRNA

A:Residues: 1-688 <RUS>

A:CROSS-references: GB:J04080; NID:g179645; PIDN:AAA51852.1; PID:g179646

R.Tosti, M.; Duponchel, C.; Meo, T.; Jullier, C.

Biochemistry 26, 8516-8524, 1987

A:title: Complete cDNA sequence of human complement C1s and close physical linkage of

A:reference number: A27381; MUID:88163522; PMID:2831944

A:Accession: A27381

A:molecule type: mRNA

A:Residues: 1-688 <TOS>

A:CROSS-references: GB:M18767; NID:g179647; PIDN:AAA51853.1; PID:g179648

R.Mackinnon, C.M.; Carter, P.E.; Smyth, S.J.; Dunbar, B.; Fothergill, J.E.

Eur. J. Biochem. 169, 547-553, 1987

A:title: Molecular cloning of cDNA for human complement component C1s. The complete a

A:reference number: S00224; MUID:88082788; PMID:3500856

A:Accession: S00224

A:molecule type: mRNA

A:Residues: 1-688 <MAC>

A:CROSS-references: EMBL:X06596; NID:g29542; PIDN:CAA29817.1; PID:g763110

A:Accession: S26732

A:molecule type: Protein

A:Residues: 16-38;168-116;170-236;246-262;265-280;282-284;287-308, 315-363;384-394;421-4

R.Tosti, M.; Duponchel, C.; Meo, T.; Couture-Tosti, E.

J. Mol. Biol. 208, 709-714, 1989

A:title: Complement genes C1r and C1s feature an intronless serine protease domain cl

A:reference number: S05634; MUID:90040704; PMID:2533964

A:Accession: S05634

A:Status: not compared with conceptual translation

A:molecule type: DNA

A:Residues: 356-513, 'G', 514-688 <TO2>

R.Carter, P.E.; Dunbar, B.; Fothergill, J.E.

Biochem. J. 215, 565-571, 1983

A:title: The serine proteinase chain of human complement component C1s. Cyanogen brom

A:reference number: A05140; MUID:84104122; PMID:6362661

A:Accession: A05140

A:molecule type: Protein

A:Residues: 438-463, 'X', 485-500;503-534;542-558;561-572, 'A', 574-601;617-623;626-644;6

R.Spycher, S.E.; Nick, H.; Rickli, E.E.

Eur. J. Biochem. 156, 49-57, 1986

A:title: Human complement component C1s. Partial sequence determination of the heavy

A:reference number: A23396; MUID:86164350; PMID:3007145

A:Accession: A23396

A:Molecule type: protein
 A:Residues: 16-61;168-219;287-293,'K',295-334;384-445 <SPY>
 R:Hess, D.; Schallier, J.; Rickli, E.E.
 Biochemistry 30, 2827-2833, 1991
 A:Title: Identification of the disulfide bonds of human complement C1s.
 A:Reference number: A38407; MUID:91175725; PMID:2007122
 A:Accession: A38407
 A:Molecule type: protein
 A:Residues: 131-134,'X',136-146,'X',148-150;155,'X',157-162;166-170,'X',172-174,'X',176-178,'X',187-402,'X',404-408;416-424,'X',426-431;547-556;592-597;617,'X',619-627,'X',629-635
 R:Thielens, N.M.; Aude, C.A.; Lacroix, M.B.; Gagnon, J.; Arlaud, G.J.
 J. Biol. Chem. 265, 14469-14475, 1990
 A:Title: Ca(2+) binding properties and Ca(2+)-dependent interactions of the isolated NH-terminus of human complement C1s.
 A:Reference number: A37820; MUID:90354439; PMID:2387866
 A:Accession: B37820
 A:Molecule type: protein
 A:Residues: 16-25,'X',203-207 <PHT>
 R:Thielens, N.M.; Van Dorsselaer, A.; Gagnon, J.; Arlaud, G.J.
 Biochemistry 29, 3570-3578, 1990
 A:Title: Chemical and functional characterization of a fragment of C1s containing the epitope recognized by monoclonal antibody 12B3.
 A:Reference number: A32672; MUID:90283368; PMID:2141278
 A:Contents: annotation: erythro-beta-hydroxyasparagine site, content
 A:Note: about half of the A chains contain erythro-beta-hydroxyasparagine
 C:Comment: This protein is a serine proteinase that combines with C1q and C1r to form C1s C2 and C4.
 C:Comment: C1s is a dimer of identical chains, each of which is activated by cleavage in C1s.
 A:Gene: GDB:C1S
 A:Cross-references: GDB:119730; OMIM:120580
 A:Map position: 12p13-12p13
 A:Introns: 291/1; 329/3; 356/1; 399/1; 424/1
 A:Note: the list of introns may be incomplete
 C:Superfamily: complement subcomponent C1r; C1r/C1s repeat homology; complement factor H C1r/C1s repeat homology; beta-hydroxyasparagine; calcium binding; complement pathway; du F-11-127/Domain: C1r/C1s repeat homology <C1R1>
 F:11-127/Domain: C1r/C1s repeat homology <C1R1>
 F:16-688/Product: complement subcomponent C1s #status experimental <MAN>
 F:16-437/Product: complement subcomponent C1s chain A (heavy chain) #status experimental
 F:135-171/Domain: EGF homology <EGF>
 F:175-287/Domain: C1r/C1s repeat homology <C1R2>
 F:294-354/Domain: complement factor H repeat homology <RH2>
 F:359-421/Domain: complement factor H repeat homology <RH2>
 F:438-688/Product: complement subcomponent C1s chain B (light chain) #status experimental
 F:438-673/Product: trypsin homology <TRY>
 F:65-83;135-147;143-156;158-171;175-202;234-251;294-341;321-354;359-403;386-421;425-549
 F:149/Modified site: erythro-beta-hydroxyasparagine (Asn) (covalent) #status experimental
 F:174;406/Blinding site: carbohydrate (Asn) (covalent) #status experimental
 F:437-438/Cleavage site: Arg-Ile (complement subcomponent C1r) #status experimental
 F:475;529;637/Active site: His, Asp, Ser #status predicted

Query Match

5.7%; Score 151; DB 1; Length 688;
 Best Local Similarity 25.3%; Pred. No. 0.0017;

Matches 46; Conservative 31; Mismatches 89; Indels 16; Gaps 5;

QY 15 FAVCAP---LRQAELGDCGHIYVSQDSGTFSTKNTPTNYVCEKITVPPGKRL 71
 DB 155 FSCPEYFLHDMKNCNGVSGDVTALIGELASPNYPPEPENSRCCEYQILKEGFOY 214
 QY 72 I--LRGDIINIESKTCAS---DYLFSSATDYGPCG--SWAPKELRLNSENTEVLFKS 125
 DB 215 VVTLREDDVDEAADSAGNCLDSLVVAGDROGPGFCGHPPLNIETKSNMIDLIFQF 274
 QY 126 GSHISGRGLTY-----ASSDHPDLITCLERGSHYEEKSKPCPACRDIADISG 178
 DB 275 DLTGKKGKMLRYHGDPMCPKEDTFNSWEPKAKYVFADVQIICLDGFEVVEGVGA 334
 QY 179 NT 180
 DB 335 TS 336

RESULT 10
 T30549

hensin - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30549
 R:Takito, J.; Hkita, C.; Al-Awqati, Q.
 J. Clin. Invest. 98, 2324-2331, 1996
 A:Title: Hensin, a new collecting duct protein involved in the in vitro plasticity of
 A:Reference number: Z20851; MUID:97096804; PMID:8941650
 A:Accession: T30549
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-1594 <TAK>
 A:Cross-references: EMBL:AF043112; MUID:94105083; PID:94105084; PIDN:AAD02242.1

Query Match
 5.6%; Score 150; DB 2; Length 1594;
 Best Local Similarity 25.6%; Pred. No. 0.0065;
 Matches 65; Conservative 34; Mismatches 105; Indels 50; Gaps 12;

QY 32 CGHIVTSQDSGTFSTKNTPTNYVCEKITVPPGKRLILRLGDIINIESKT-CASDYL 90
 DB 951 CGGFLVNA-TGSPSSPSPYPYPNACLWEIAPVSGYLINLFGSQLRLEQHSVCNEDY 1009
 QY 91 -LSSARDQ--YPRYCSMAVPRKELRLNSENTEVLFKSSGSHISGRGLTYAS----- 140
 DB 1010 EIRPDGSDSSLLKRICNDG--QIFTTSSNRMTVLFPSDLSVONTGLFVANSFPDASL 1067
 QY 141 ---SDHEDLITCLERGSHYEEKSKPCPACRDIADISGNTKDG-YRDTSLCKRAIH 196
 DB 1068 RLVSGNSYACAGRAVEIYHGRMGTYC-----DSWDVQDAQVYCRQDGGAVS 1118
 QY 197 A-----GIIT-DEL---GGHINLQSKGISHYEGLLNGVLSRRGSLSEKRLFT 242
 DB 1119 APGAYVSGSSGGLTLDVNCSTGTEALIMQCRSQSWP-----SHNGHEDASVIC 1169
 QY 243 TPGMNTITVAIPSY 256
 DB 1170 TGNVGTTSVSPNT 1183

RESULT 11
 CHIDRB
 complement subcomponent C1RBAR.GIF (EC 3.4.21.41) precursor [validated] - human
 C:Species: Homo sapiens (man)
 C:Date: 15-Nov-1984 #sequence_revision 30-Jun-1991 #text_change 03-Jun-2002
 R:Leysen, A.; A24170; A29768; A29769; S02422; A00916; A37820; S68830
 C:Leysen, S.P.; Kurachi, K.; Sakaribassen, K.S.; Davie, E.W.
 Biochemistry 25, 4855-4863, 1986
 A:Title: Nucleotide sequence of the cDNA coding for human complement C1r.
 A:Reference number: A24170; MUID:87026566; PMID:3021205
 A:Accession: A24170
 A:Molecule type: mRNA
 A:Residues: 1-705 <LEY>
 A:Cross-references: GB:M4058; NID:q179643; PIDN:AAA51851.1; PID:q179644
 R:Journet, A.; Tosi, M.
 Biochem. J. 240, 783-787, 1986
 A:Title: Cloning and sequencing of full-length cDNA encoding the precursor of human
 A:Reference number: A29768; MUID:87156625; PMID:3030286
 A:Accession: A29768
 A:Molecule type: mRNA
 A:Residues: 1-151,'L',153-705 <JOU>
 A:Cross-references: GB:X04701; NID:g29538; PIDN:CAA28407.1; PID:g29539
 R:Arlaud, G.J.; Willis, A.C.; Gagnon, J.
 Biochem. J. 241, 711-720, 1987
 A:Title: Complete amino acid sequence of the A chain of human complement-classical-pa
 A:Reference number: A29769; MUID:87241248; PMID:3036070
 A:Accession: A29769
 A:Molecule type: protein
 A:Residues: 18-166,'X',168-463 <ARL>
 A:Note: 152-Leu was also found
 R:Arlaud, G.J.; Van Dorsselaer, A.; Bell, A.; Mancini, M.; Aude, C.; Gagnon, J.
 FEBS Lett. 222, 129-134, 1987
 A:Title: Identification of erythro-beta-hydroxyasparagine in the EGF-like domain of h

A: Molecule type: mRNA
A: Residues: 1-991 <RES>
A: Cross-references: GB:124755; NID:9439606; PIDN:AAA37306.1; PID:9439607
C: Genes: 1
A: Gene: Bmp-1
C: Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology; F:135-326/Domain: astacin homology <AST>
C: Keywords: hydrolase; metalloproteinase; zinc
F: 556-592/Domain: EGF homology <EGF1>
F: 596-705/Domain: C1r/C1s repeat homology <C1R>
F: 712-747/Domain: EGF homology <EG2>
F: 218,222,228,277/Binding site: zinc (His, His, His, Tyr) #status predicted
F: 219/Active site: Glu #status predicted

Query Match 5.3%; Score 142.5; DB 2; Length 991;
Best Local Similarity 33.0%; Pred. No. 0.013;
Matches 38; Conservative 22; Mismatches 48; Indels 7; Gaps 4;

QY 31 GCGHIVTSQDSGWTSMKNYPGYNYVCEKIIIVPKGRLLRLGLDNIIESK-TCASDY 89
DB 751 GCEHRTVS-TSGRTSPNWDKYPKREKCEWAIISSTPGHRYKLTFFVEMDIESQECAYDH 809
QY 90 LFFSSATDQYGR---YCGSWAVPKELRLNSNEYVLEKSGSHISGRGFLITVAS 140
DB 810 LEVFDGRDAKAPVIGRFCS-KRPEVLAIGNRFLEFSDNSVQRKGFQASHST 863

RESULT 14

JC7629
membrane-type frizzled-related protein - human
C: Species: Homo sapiens (man)
C: Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C: Accession: JC7629
R: Katoh, M.
Biochem. Biophys. Res. Commun. 282, 116-123, 2001
A: Title: Molecular cloning and characterization of MFRP, a novel gene encoding a membrane-type frizzled-related protein
A: Reference number: JC7629; MUID:21164708; PMID:11263980
A: Molecule type: mRNA
A: Accession: JC7629
A: Residues: 1-579 <KAT>
A: Cross-references: DDB:AB055505
C: Comment: This protein, which plays key roles in medulla oblongata as a regulator of thubulin domains.
C: Genes: 1
A: Gene: mfrp
A: Map position: 11q23
C: Keywords: transmembrane protein

Query Match 5.3%; Score 142; DB 2; Length 579;
Best Local Similarity 23.7%; Pred. No. 0.0064;
Matches 58; Conservative 37; Mismatches 88; Indels 62; Gaps 10;

QY 31 GCGHIVTSQDSGWTSMKNYPGYNYVCEKIIIVPKGRLLRLGLDNIIESK-TCASDY 89
DB 300 GCGGNLTGL-QGTFSTPSTYLAQVPHOLCTWHISVPAGHSITELQPHNPSLEADECKFDY 358
QY 90 L-----LSSATDQYGRPGYNYVCEKIIIVPKGRLLRLGLDNIIESK-TCASDY 142
DB 359 VEYETSSGAFSLGRFCGA-EPPPHLVSHHSLAVLFRDHISSGFSATYLAFNAT 417
QY 143 ----HPLITCLERGSYFEKYSKFCPCAGCDIAG--DISGNTKDGTRDT----- 187
DB 418 ENPGSPSELSCQ-----AGCGKGVQMMCDMWRDCTGSDSDNCSGPLFPP 462
QY 188 SLCKNA-----IHAGITDE-----LGGHINLQSKGISHEGLLAN 225
DB 463 ELACEPQVEMCGLSYNTTAFFNIVGMITIQEEVVELSGYSLTSLPCYQHRRLLCG 522
QY 226 GVLNR 230
DB 523 LTVPR 527

RESULT 15

T42721
CRP-ductin-alpha precursor - mouse
C: Species: Mus musculus (house mouse)
C: Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 23-Mar-2001
C: Accession: T42721
R: Cheng, H.; Bjerknes, M.; Chen, H.
Anat. Rec. 244, 327-343, 1996
A: Title: CRP-ductin: a gene expressed in intestinal crypts and in pancreatic and hepa
A: Reference number: 222241; MUID:96362470; PMID:8742698
A: Accession: T42721
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: mRNA
A: Residues: 1-2083 <CHE>
A: Cross-references: EMBL:U37438; NID:91276646; PID:91276647; PIDN:AAC52505.1
A: Experimental source: strain BALB/c; jejunal epithelial cells
C: Keywords: transmembrane protein
F: 1-28/Domain: signal sequence #status predicted <SIG>
F: 29-2083/Product: CRP-ductin-alpha #status predicted <MAT>

Query Match 5.3%; Score 142; DB 2; Length 2083;
Best Local Similarity 24.1%; Pred. No. 0.04;
Matches 70; Conservative 39; Mismatches 91; Indels 90; Gaps 16;

QY 32 GCHIVTSQDSGWTSMKNYPGYNYVCEKIIIVPKGRLLRLGLDNIIESKTCASDY- 90
DB 1631 GCGFLT-QPSGQFSSPFYPGNYPNMARCLMNIIEVPPNRYVVRDQLLE-RGCSYDYE 1688
QY 91 LFFSSATDQYGRPGYNYVCEKIIIVPKGRLLRLGLDNIIESKTCASDY- 143
DB 1689 IFD-----GPHSSPLIARVCDLSGFSFTSNFMSIRFTDHSIYARGP----- 1733
QY 144 PDLITCLERGSYFEKYSKFCPCAGCDIAGDISGNTKDGTRDYSLCKNAIHAGITDE 203
DB 1734 -----QAHY-----SDPDNT-----TNLQCS----- 1752
QY 204 LGHI-----NLQSGISHYEGHLANGVLSRH--GSLSEKRLFTTPGMNITTV---- 251
DB 1753 --NHQASVSRSYLSQMSYARDLVIPGMNSYHCQPOITQREVIFITPYGCGTIRKQAD 1810
QY 252 --ALPSYFIALLTGWIGIFAIKRRKKGPVYSAD-AQRTGCKQDIX 298
DB 1811 NETINYSFLRAVVSN-GIT---KRRKDLNIHSCRLQNT--WVNTMY 1853

Search completed: May 15, 2003, 13:17:00
Job time : 23.454 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 13.6281 Seconds

(Without alignments)
1530.854 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGPFVLLAFVACAP.....YSAPRNGLAFLNQTATALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181.5	6.8	686	1	MAS2_HUMAN
2	158.5	5.9	695	1	CASP_MESAU
3	157.5	5.9	704	1	CRAR_MOUSE
4	151.5	5.7	699	1	NRPI_XENLA
5	151.5	5.7	688	1	CRAR_HUMAN
6	147	5.5	705	1	CIR_HUMAN
7	145	5.4	550	1	COCH_HUMAN
8	143.5	5.4	923	1	NRPI_MOUSE
9	142.5	5.3	991	1	BMP1_MOUSE
10	141.5	5.3	986	1	BMP1_HUMAN
11	141.5	5.3	514	1	UVS2_XENLA
12	141	5.3	639	1	BMP1_STRPU
13	140.5	5.3	468	1	PCOI_RAT
14	139	5.2	1022	1	TLID_BRARE
15	135	5.1	449	1	PCOI_HUMAN
16	133.5	5.0	468	1	NRPI_MOUSE
17	132.5	4.9	931	1	NRPI_HUMAN
18	132	4.9	547	1	COCH_CHICK
19	131.5	4.9	855	1	STI4_HUMAN
20	131.5	4.9	922	1	NRPI_RAT
21	131.5	4.9	925	1	NRPI_RAT
22	130	4.9	552	1	COCH_MOUSE
23	129.5	4.8	707	1	NRPI_MOUSE
24	127.5	4.8	931	1	NRPI_XENLA
25	127	4.8	855	1	STI4_MOUSE
26	125	4.7	1019	1	LFC_MOUSE
27	125	4.6	1057	1	LFC_DROME
28	122	4.6	616	1	SPAN_STRPU
29	121.5	4.5	923	1	NRPI_HUMAN
30	121	4.5	1019	1	LFC_CARRO
31	120.5	4.5	597	1	BPIO_PARLI
32	119.5	4.5	914	1	NRPI_CHICK
33					P79795 gallus gall

34	116	4.3	275	1	TSG6_MOUSE	O08859 mus musculus
35	115	4.3	277	1	TSG6_HUMAN	P98066 homo sapien
36	113.5	4.2	1064	1	FBP1_STRPU	P10079 strongyloce
37	110	4.1	276	1	TSG6_RABIT	P98065 oryctolagus
38	110	4.1	1369	1	MSH5_CAEEL	Q19272 caenorhabdi
39	105.5	3.9	1493	1	NEO1_MOUSE	P97798 mus musculus
40	105	3.9	620	1	EXTN_TOBAC	P13963 nicotiana t
41	105	3.9	968	1	BCAL_RAT	O63767 rattus norv
42	104.5	3.9	570	1	FBP3_STRPU	P49013 strongyloce
43	104.5	3.9	1069	1	ENTR_MOUSE	P97435 mus musculus
44	104.5	3.9	1461	1	NEO1_HUMAN	Q92859 homo sapien
45	104	3.9	310	1	ASTL_CORJIA	P42662 coturnix co

ALIGNMENTS

RESULT 1
MAS2_HUMAN STANDARD; PRT; 686 AA.
ID MAS2_HUMAN
AC 000187; 075754; 09Y270; 09B2H0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mannan-binding lectin serine protease 2 precursor (EC 3.4.21.-)
DE (Mannose-binding protein associated serine protease 2) (MASP-2)
GN MASP2
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97242412; PubMed=9087411;
RA Thiel S., Jensen T.V., Stover C.M., Schwaible W.J., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "A second serine protease associated with mannan-binding lectin that
RT activates complement.";
RL Nature 386:506-510(1997).
[2]
RN SEQUENCE FROM N.A.
RX Thiel S., Vorup-Jensen T., Stover C.M., Schwaible W.J., Laursen S.B.,
RA Poulsen K., Willis A.C., Eggleston P., Hansen S., Holmskov U.,
RA Reid K.B.M., Jensenius J.C.;
RT "Identification and characterization of a novel protein of the human
RT complement system, mannan-binding lectin-associated serine protease-2
RT (MASP-2).";
RN Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=99192764; PubMed=10092804;
RA Stover C.M., Thiel S., Thelen M., Lynch N.J., Vorup-Jensen T.,
RA Jensenius J.C., Schwaible W.J.;
RT "Two constituents of the initiation complex of the mannan-binding
RT lectin activation pathway of complement are encoded by a single
RT structural gene.";
RL J. Immunol. 162:3481-3490(1999).
[4]
RN SEQUENCE FROM N.A.
RX Park D., Kim B., Baek K., Yoon J.;
RT "Structure of human MASP-2 gene.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN FUNCTION: TRYPSIN PROTEASE THAT PROBABLY PLAYS AN IMPORTANT ROLE
RN IN THE INITIATION OF THE MBL COMPLEMENT ACTIVATION PATHWAY. AFTER
RN ACTIVATION IT CLEAVES C4 GENERATING C4A AND C4B.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC - SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 1 BGF-LIKE DOMAIN.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X16160; CAA34286.1; -
 DR PIR: S05008; S05008.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.193; -
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB domain.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF000084; sushi_1.
 DR Pfam: PF000084; sushi_2.
 DR Pfam: PF00089; trypsin_1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; FALSE_NEG.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS02040; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR Hydroxylase: Serine protease; Calcium-binding; Extracellular matrix;
 KW Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation; Glycoprotein;
 KM Signal; EGF-like domain; Repeat.
 FT SIGNAL 1 21
 FT CHAIN 22 695
 FT CHAIN 22 444
 FT CHAIN 445 695
 FT DOMAIN 22 124
 FT DOMAIN 22 136
 FT DOMAIN 181 296
 FT DOMAIN 297 349
 FT DOMAIN 352 415
 FT DOMAIN 445 695
 FT DOMAIN 137 178
 FT MOD_RES 155 155
 FT ACT_SITE 482 482
 FT ACT_SITE 536 536
 FT ACT_SITE 638 638
 FT DISULFID 141 153
 FT DISULFID 149 162
 FT DISULFID 164 177
 FT CARBOHYD 180 180
 FT CARBOHYD 413 413
 SQ SEQUENCE 695 AA; 77493 MW; E924F7E6340700D0 CRC64;
 Query Match 5.98; Score 158.5; DB 1; Length 695;
 Best Local Similarity 25.88; Pred. No. 0.00015;
 Matches 48; Conservative 37; Mismatches 76; Indels 25; Gaps 7;

QY 175 DISGNT 180
 DB 337 NV-GST 341

RESULT 3
 ID CRAR_MOUSE STANDARD; PRT; 704 AA.
 AC P98064;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement-activating component of Ra-reactive factor precursor
 DE (EC 3.4.21.-) (Ra-reactive factor serine protease p100) (RaRF)
 DE (Mannan-binding lectin serine protease 1).
 GN MASP1 OR CRARF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX MEDLINE=94179811; PubMed=8133044;
 RA Takayama Y., Takada F., Takahashi A., Kawakami M.;
 RT "A 100-kDa protein in the C4-activating component of Ra-reactive
 RT factor is a new serine protease having module organization similar to
 RT C1r and C1s."
 RL J. Immunol. 152:2308-2316(1994).
 RN [2]
 RP SEQUENCE OF 465-704 FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=BALB/C; TISSUE=Liver;
 RX MEDLINE=93176166; PubMed=8439319;
 RA Takahashi A., Takayama Y., Hattuse H., Kawakami M.;
 RT "Presence of a serine protease in the complement-activating component
 RT of the complement-dependent bactericidal factor, RaRF, in mouse
 RT serum."
 RL Biochem. Biophys. Res. Commun. 190:681-687(1993).
 RN [2]
 RP FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF
 WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
 CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
 CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
 C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
 CC -1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
 (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
 CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (29)
 CC LINKED BY A DISULFIDE BOND.
 CC -1- TISSUE SPECIFICITY: LIVER.
 CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00431; CUB; 2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; TRYP_SPE; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS00240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 DR Hydrolase: Complement pathway: Serine protease; Protease;
 KW Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.
 FT SIGNAL 1 24.
 FT CHAIN 25 704
 FT CHAIN 25 453
 FT CHAIN 454 704
 FT CHAIN 25 143
 FT CHAIN 144 187
 FT CHAIN 190 302
 FT CHAIN 305 368
 FT CHAIN 371 438
 FT CHAIN 454 704
 FT CHAIN 495 495
 FT CHAIN 557 557
 FT CHAIN 651 651
 FT MOD_RES 164 164
 FT DISULFID 78 96
 FT DISULFID 148 162
 FT DISULFID 158 171
 FT DISULFID 173 186
 FT DISULFID 190 217
 FT DISULFID 247 265
 FT DISULFID 306 354
 FT DISULFID 372 419
 FT DISULFID 402 437
 FT DISULFID 441 577
 FT DISULFID 619 636
 FT DISULFID 647 677
 FT CARBOHYD 54 54
 FT CARBOHYD 183 183
 FT CARBOHYD 390 390
 FT CARBOHYD 412 412
 FT SEQUENCE 704 AA; 79895 MM; 71P44F3012D2C67F CRC64;
 Query Match 5.9%; Score 157.5; DB 1; Length 704;
 Best Local Similarity 24.1%; Pred. No. 0.00019;
 Matches 58; Conservative 53; Mismatches 93; Indels 37; Gaps 12;

DB 428 K 428
 RESULT 4
 NRPI_XENIA STANDARD: PRT; 928 AA.
 ID NRPI_XENIA
 AC P28824;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neupilin-1 precursor (A5 protein) (A5 antigen).
 OS Xenopus laevis (African clawed frog).
 OC Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91337458; PubMed=1908252;
 RA Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;
 RT "The A5 antigen, a candidate for the neuronal recognition molecule,
 RL has homologues to complement components and coagulation factors.";
 RL Neuron 7:295-307(1991).
 CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CEREAL NEURONAL CIRCUITS AND IN ORGANOREGULATION OF SEMAPHORINS (BY
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
 CC BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
 CC SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
 CC NEURONS.
 CC -1- SIMILARITY: BELONGS TO THE NEUPILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D10467; BA01260.1; -.
 DR HSSP: P12259; ICZF.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR00421; FA58_C.
 DR InterPro: IPR00998; MAM_domain.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR PRINTS: PR00020; MAMDOMAIN.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00231; FA58C; 2.
 DR SMART: SM00137; MAM; 1.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01285; FA58C_1; 2.
 DR PROSITE: PS01286; FA58C_2; 2.
 DR PROSITE: PS50060; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurex; Signal; Repeat; Receptor;
 KW Antigen.
 FT SIGNAL 1 21
 FT CHAIN 22 928
 FT CHAIN 22 860
 FT TRANSMEM 861 883
 FT DOMAIN 884 928
 FT DOMAIN 27 141
 FT DOMAIN 147 265
 FT DOMAIN CUB 1.
 FT DOMAIN CUB 2.

[1]
RP SEQUENCE FROM N.A.
RC TISSUE-Liver;
RX MEDLINE-94059062; PubMed-8240317;
RA Takada F., Takayama Y., Hatsuue H., Kawakami M.;
RT "A new member of the C1s family of complement proteins found in a
RL bactericidal factor. Ra-reactive factor. In human serum.";
RN Biochem. Biophys. Res. Commun. 196:1003-1009(1993).
RP [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Fetal liver;
RX MEDLINE-94289349; PubMed-8018603;
RA Sato T., Endo Y., Matsushita M., Fujita T.;
RT "Molecular characterization of a novel serine protease involved in
RL activation of the complement system by mannose-binding protein.";
RN Int. Immunol. 6:665-669(1994).
RP [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-97079701; PubMed-89221412;
RA Endo Y., Sato T., Matsushita M., Fujita T.;
RT "Exon structure of the gene encoding the human mannose-binding
RL protein-associated serine protease light chain: comparison with
RN complement C1r and C1s genes.";
RP Int. Immunol. 8:1355-1358(1996).
RP [4]
RP SEQUENCE FROM N.A.
RC MEDLINE-99402590; PubMed-10475605;
RX Takayama Y., Takada F., Nowatari M., Kawakami M., Matsura N.;
RA "Gene structure of the P100 serine-protease component of the human Ra-
RT reactive factor.";
RN Mol. Immunol. 36:505-514(1999).
RP [5]
RP FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF
CC WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
CC CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
CC CASCADE BY CLEAVING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
CC C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
CC -1 SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
CC (CARAF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
CC CARAF IS AN HEMEROPOLYMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)
CC LINKED BY A DISULFIDE BOND.
CC -1 DOMAIN: CARAF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1 SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1 SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1 SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: D17525; BAA04477.1; -
DR EMBL: D28593; BAA05928.1; -
DR EMBL: D61695; BAA34864.1; -
DR EMBL: AB010822; BAA34864.1; JOINED.
DR EMBL: AB010813; BAA34864.1; JOINED.
DR EMBL: AB010814; BAA34864.1; JOINED.
DR EMBL: AB010815; BAA34864.1; JOINED.
DR EMBL: AB010816; BAA34864.1; JOINED.
DR EMBL: AB010817; BAA34864.1; JOINED.
DR EMBL: AB010818; BAA34864.1; JOINED.
DR EMBL: AB010819; BAA34864.1; JOINED.
DR EMBL: AB010820; BAA34864.1; JOINED.
DR EMBL: AB010821; BAA34864.1; JOINED.
DR EMBL: D61690; BAA34864.1; JOINED.
DR EMBL: D61691; BAA34864.1; JOINED.
DR EMBL: D61692; BAA34864.1; JOINED.
DR EMBL: D61693; BAA34864.1; JOINED.
DR EMBL: D61694; BAA34864.1; JOINED.

RC TISSUE-Peripheral blood leukocytes;
RX MEDLINE-99008558; Pubmed-9794427;
RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.,
RT "Two lineages of mannose-binding lectin-associated serine protease
RT (MASP) in vertebrates.";
RL J. Immunol. 161:4924-4930(1998).
RN [5]
RP SEQUENCE OF 291-688 FROM N.A.
RX MEDLINE-90040704; Pubmed-2553994;
RA Tosi M., Duponchel C., Moio T., Couture-Tosi E.,
RT "Complement genes C1r and C1s feature an intronless serine protease
RT domain closely related to haptoglobin.";
RL J. Mol. Biol. 208:709-714(1989).
RN [6]
RP SEQUENCE OF 16-61; 168-219; 287-334 AND 384-445.
RX MEDLINE-86164350; Pubmed-3007145;
RA Spycher S.E., Nick H., Rickli E.E.,
RT "Human complement component C1s. Partial sequence determination of
RT the heavy chain and identification of the peptide bond cleaved during
RT activation.";
RL Eur. J. Biochem. 156:49-57(1986).
RN [7]
RP SEQUENCE OF 438-500; 503-534; 542-601; 617-623 AND 626-656.
RX MEDLINE-84104122; Pubmed-6332651;
RA Carter P.E., Dunbar B., Fothergill J.E.,
RT "The serine proteinase chain of human complement component C1s.
RT Cysteine bromide cleavage and N-terminal sequences of the
RT fragments.";
RL Biochem. J. 215:565-571(1983).
RN [8]
RP PARTIAL SEQUENCE.
RX TISSUE-Plasma;
RX MEDLINE-91308095; Pubmed-18544725;
RA Iilly C., Thielems N.M., Gagnon J., Arlaud G.J.,
RT "Effect of laccoperoxidase-catalyzed iodination on the Ca(2+)-
RT dependent interactions of human C1s. Location of the iodination
RT sites.";
RL Biochemistry 30:7135-7141(1991).
RN [9]
RP MEDLINE-911575725; Pubmed-20071122;
RX Hess D., Schaller J., Rickli E.E.,
RA "Identification of the disulfide bonds of human complement C1s.";
RL Biochemistry 30:2827-2833(1991).
RN [10]
RP PARTIAL SEQUENCE, AND 3D-STRUCTURE MODELING OF CATALYTIC DOMAIN.
RX MEDLINE-95298736; Pubmed-7719774;
RA Rossi V., Gaboriaud C., Lacroix M., Ulrich J., Fontecilla-Camps J.C.,
RA Gagnon J., Arlaud G.J.,
RT "Structure of the catalytic region of human complement protease C1s:
RT study by chemical cross-linking and three-dimensional homology
RT modeling.";
RL Biochemistry 34:7311-7321(1995).
RN [11]
RP DISEASE.
RX Pubmed-11390518;
RA Dragon-Durey M.-A., Quartier P., Fremiaux-Bacchi V., Blouin J.,
RT de Baracze C., Pileur A.-M., Weiss L., Fridman W.-H.,
RT "Molecular basis of a selective C1s deficiency associated with early
RT onset multiple autoimmune diseases.";
RL J. Immunol. 166:7612-7616(2001).
RN [12]
RP FUNCTION. C1S B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1Q
CC AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
CC OF THE COMPLEMENT SYSTEM. C1R ACTIVATES C1S SO THAT IT CAN, IN
CC TURN, ACTIVATE C2 AND C4.
CC -1- CATALYTIC ACTIVITY: CLEAVES COMPONENT C4 TO C4A AND C4B, AND
CC COMPONENT C2 TO C2A AND C2B.
CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
CC AND S IN THE MOLAR RATION OF 1:2:2. ACTIVATED C1S IS AN DISULFIDE
CC LINKED HETERODIMER OF AN HEAVY CHAIN AND A LIGHT CHAIN.
CC -1- DISEASE: Defects in C1S are the cause of selective C1s deficiency
CC that is associated with early onset multiple autoimmune diseases.

CC	-1-	SIMILARITY:	BELONGS TO PEPTIDASE FAMILY S1.
CC	-1-	SIMILARITY:	CONTAINS 2 SUSI1 (SCP) DOMAINS.
CC	-1-	SIMILARITY:	CONTAINS 2 CUB DOMAINS.
CC	-1-	SIMILARITY:	CONTAINS 1 EGF-LIKE DOMAIN.
CC	-1-	SIMILARITY:	CONTAINS 1 EGF-LIKE DOMAIN.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and thus statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).		
CC	or send an email to	license@isb-sib.ch)	.
DR	EMBL; X06596;	CAB2887.1;	--
DR	EMBL; J04080;	AAS1852.1;	--
DR	EMBL; M18767;	AAS1853.1;	--
DR	EMBL; AB09076;	BAA6864.1;	--
DR	PfR; A40456;	CIHOS.	--
DR	PfR; S00224;	S00224.	--
DR	MEROPS; S01.193;	--	--
DR	SWISS-2DPAGE; P09871;	HUMAN.	--
DR	Genev; HGNC:1247;	CIS.	--
DR	MIM; 120580;	--	--
DR	InterPro; IPR000152;	Asx_hydroxy1.	--
DR	InterPro; IPR000859;	CUB_domain.	--
DR	InterPro; IPR001314;	Chymotrypsin.	--
DR	InterPro; IPR000561;	EGF-like.	--
DR	InterPro; IPR001881;	EGF_Ca.	--
DR	InterPro; IPR001254;	Set_protase_Try.	--
DR	InterPro; IPR00436;	Sushl_SCR CCP.	--
DR	Pfam; PF00008;	Egf_1.	--
DR	Pfam; PF00084;	sushl_2.	--
DR	Pfam; PF00089;	tryptsin; 1.	--
DR	Pfam; PF00431;	CUB; 2.	--
DR	PRINTS; PR00722;	CHYMOTRYPSIN.	--
DR	SMART; SM00032;	CCP; 2.	--
DR	SMART; SM00042;	CUB; 2.	--
DR	SMART; SM00179;	EGF_CA; 1.	--
DR	SMART; SM00020;	Tryp_Spc; 1.	--
DR	PROSITE; PS00010;	ASX_HYDROXYL; 1.	--
DR	PROSITE; PS01180;	CUB; 2.	--
DR	PROSITE; PS00022;	EGF_1; FALSE_NEG.	--
DR	PROSITE; PS01186;	EGF_2; FALSE_NEG.	--
DR	PROSITE; PS01187;	EGF_CA; 1.	--
DR	PROSITE; PS050240;	TRYP SIN_DOM; 1.	--
DR	PROSITE; PS00134;	TRYPSIN_HIS; FALSE_NEG.	--
DR	PROSITE; PS00135;	TRYPSIN_SER; 1.	--
KW	Complement pathway; Plasma;	Glycoprotein; Serine protease; Hydrolase;	--
KW	Hydroxylation; Sushl; Repeat;	Signal; Egf-like domain;	--
KM	Calcium-binding.	--	--
FT	SIGNAL	1	15
FT	CHAIN	16	437
FT	DOMAIN	438	688
FT	DOMAIN	16	130
FT	DOMAIN	131	172
FT	DOMAIN	175	290
FT	DOMAIN	293	355
FT	DOMAIN	358	422
FT	DOMAIN	438	688
FT	ACT_SITE	475	475
FT	ACT_SITE	529	529
FT	MOD_RES	632	632
FT	CARBONHD	149	149
FT	CARBONHD	174	174
FT	DISELFED	406	406
FT	DISELFED	65	83
FT	DISELFED	135	147
FT	DISELFED	143	156
FT	DISELFED	158	171
FT	DISELFED	175	202
FT	DISELFED	234	251
FT	DISELFED	294	341
FT	DISELFED	321	354

```

FT DISULFID 359 403
FT DISULFID 386 421
FT DISULFID 425 549
FT DISULFID 595 618
FT DISULFID 628 659
FT CONFLICT 294 294
FT CONFLICT 513 513
FT CONFLICT 573 573
FT CONFLICT 645 646
SQ SEQUENCE 688 AA; 76684 MM; 85522647A4C47205 CRC64;

Query Match 5.7%; Score 151; DB 1: Length 688;
Best Local Similarity 25.3%; Pred. No. 0.00056;
Matches 46; Conservative 31; Mismatches 89; Indels 16; Gaps 5;

QY 15 FAVCAP--LRLQAEILDGCGHIVTSODSGTMTSKNPGYPTNYVCEKITVPRGRRL 71
DQ 155 FCSCEPREFLNDKMKNCVNCSDVFTALIGELIAPNPKYPENSKCEYQIRLEKGRQV 214
QY 72 I-LRLGDLNIESKTCAS--DYLLFSSATDQYGEYCG-SMAVPEKRLNNSNVTLPKS 125
DQ 215 VVTLRREFDVEADSDAGNCNLSLTFVAGDROFGPYCGHCPGPLNITKSMALDIIFQT 274
QY 126 GSHISGRFLTY-----ASSDHPDLITCLERSHFEERYSKFCFAGCRDINDISG 178
DQ 275 DLTSQKRWKMLRIGHDPMPCREKEDTPNSWEPAPKAKYFRDVQITCLDGEVEGRVGA 334
QY 179 NT 180
DQ 335 TS 336

RESULT 7
CIR_HUMAN STANDARD; PRT; 705 AA.
ID CIR_HUMAN
AC P00736;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement C1r component precursor (EC 3.4.21.41).
GN CIR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE-87026566; PubMed-3021205;
RA Leytus S.P., Kurachi K., Sakariassen K.S., Davie E.W.;
RT "Nucleotide sequence of the cDNA coding for human complement C1r.";
RL Biochemistry 25:485-486(1986).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE-87156625; PubMed-3030286;
RA Journet A., Tosi M.;
RT "Cloning and sequencing of full-length cDNA encoding the precursor of
human complement component C1r.";
RL Biochem. J. 240:783-787(1986).
RN 13
RP SEQUENCE OF 18-463.
RX MEDLINE-87241248; PubMed-3036070;
RA Arlaud G.J., Willis A.C., Gagnon J.;
RT "Complete amino acid sequence of the A chain of human complement-
classical pathway enzyme C1r.";
RL Biochem. J. 241:711-720(1987).
RN 14
RP SEQUENCE OF 464-705.
RX MEDLINE-83204782; PubMed-6303394;
RA Arlaud G.J., Gagnon J.;
RT "Complete amino acid sequence of the catalytic chain of human
complement subcomponent C1-r.";
RL Biochemistry 22:1758-1764(1983).
RN 15

```

```

RP SEQUENCE OF 152-186, AND HYDROXYLATION.
RX MEDLINE-88005128; PubMed-2820791;
RA Arlaud G.J., van Dorsselaer A., Bell A., Mancini M., Aude C.,
RA Gagnon J.;
RT "Identification of erythro-beta-hydroxyasparagine in the EGF-like
domain of human C1r.";
RL FEBS Lett. 222:129-134(1987).
RN 16
RP SEQUENCE OF 133-137; 187-211 AND 609-613, AND PHOSPHORYLATION.
RX MEDLINE-96221263; PubMed-8635594;
RA Pelloux S., Thielen N.M., Hudry-Clergeon G., Petillot Y., Filhol O.,
RA Arlaud G.J.;
RT "Identification of a cryptic protein kinase CK2 phosphorylation site
in human complement protease C1r, and its use to probe intramolecular
interaction.";
RL FEBS Lett. 386:15-20(1996).
RN 17
RP STRUCTURE BY NMR OF 140-192.
RX MEDLINE-98138432; PubMed-9477945;
RA Bersch B., Hernandez J.-F., Marion D., Arlaud G.J.;
RT "Solution structure of the epidermal growth factor (EGF)-like module
of human complement protease C1r, an atypical member of the EGF
family.";
RL Biochemistry 37:1204-1214(1998).
CC -1- FUNCTION: C1R B CHAIN IS A SERINE PROTEASE THAT COMBINES WITH C1Q
AND C1S TO FORM C1, THE FIRST COMPONENT OF THE CLASSICAL PATHWAY
OF THE COMPLEMENT SYSTEM.
CC -1- CATALYTIC ACTIVITY: ACTIVATES C1S BY PROTEOLYTIC CLEAVAGE SO
THAT IT CAN, IN TURN, ACTIVATE C2 AND C4.
CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R
AND S IN THE MOLAR RATION OF 1:2:2. C1R IS A DIMER OF IDENTICAL
CHAINS, EACH OF WHICH IS ACTIVATED BY CLEAVAGE INTO TWO CHAINS, A
AND B, CONNECTED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by, and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04701; CA28407.1; -
DR EMBL; M14058; AAA51851.1; -
DR PIR; A24170; C1HURB
DR PDB; 1APQ; 17-SEP-97.
DR MEROPS; S01.192; -.
DR Genew; HGNC:1246; C1R.
DR MIM; 216950; -.
DR InterPro; IPR000152; Asx_hydroxyl-
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 2.
DR Pfam; PF00089; Lysp1n; 1.
DR Pfam; PF00431; CUB; 2.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00032; CCP; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.

```

DR PROSITE; PS50240; TRYPsin.DOM. 1.
 DR PROSITE; PS00134; TRYPsin.HIS; FALSE_NEG.
 DR PROSITE; PS00135; TRYPsin.SER. 1.
 KW Complement pathway; Plasma: Glycoprotein; Serine protease; Hydrolase;
 KW Hydroxylation; Phosphorylation; Sush1; Repeat; Signal;
 KW EGF-like domain; 3d-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 463 COMPLEMENT C1R HEAVY CHAIN.
 FT CHAIN 18 464 705 COMPLEMENT C1R LIGHT CHAIN.
 FT DOMAIN 142 190 CUB 1.
 FT DOMAIN 193 305 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 308 372 CUB 2.
 FT DOMAIN 375 448 SUSH1 1.
 FT DOMAIN 464 705 SERINE PROTEASE.
 FT ACT SITE 502 552 CHARGE RELAY SYSTEM.
 FT ACT SITE 557 557 CHARGE RELAY SYSTEM.
 FT ACT SITE 654 654 CHARGE RELAY SYSTEM.
 FT MOD_RES 167 167 HYDROXYLATION.
 FT MOD_RES 206 206 PHOSPHORYLATION (BY CK2).
 FT CARBOHD 125 125 N-LINKED (GLCNAC. . .).
 FT CARBOHD 221 221 N-LINKED (GLCNAC. . .).
 FT CARBOHD 514 514 N-LINKED (GLCNAC. . .).
 FT CARBOHD 581 581 N-LINKED (GLCNAC. . .).
 FT DISULFD 71 89 PROBABLE.
 FT DISULFD 146 165
 FT DISULFD 161 174
 FT DISULFD 176 189
 FT DISULFD 193 220 PROBABLE.
 FT DISULFD 250 268 PROBABLE.
 FT DISULFD 309 358 PROBABLE.
 FT DISULFD 338 371 PROBABLE.
 FT DISULFD 376 429 PROBABLE.
 FT DISULFD 406 447 PROBABLE.
 FT DISULFD 451 577 INTERCHAIN (PROBABLE).
 FT DISULFD 620 639 POTENTIAL.
 FT DISULFD 650 680
 FT CONFLICT 152 152
 SO SEQUENCE 705 AA; 80173 MW; 5CBCC0201061463 CRC64;
 Query Match 5.5%; Score 147; DB 1; Length 705;
 Best Local Similarity 27.7%; Pred. No. 0.0012;
 Matches 52; Conservative 30; Mismatches 84; Indels 22; Gaps 7;

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cochlea;
 RX MEDLINE=98110569; PubMed=9441737;
 RA Robertson N.G., Skvorak A.B., Yin Y., Weremowicz S., Johnson K.R.,
 RA Kovach K.A., Batley J.F., Bieker F.R., Morton C.C.;
 RA "Mapping and characterization of a novel cochlear gene in human and in
 mouse: a positional candidate gene for a deafness disorder, DFNA9.",
 RA Genomics 46:345-354(1997).
 RL [2]
 RN VARIANTS DFNA9 GLY-66; GLU-88 AND ARG-117.
 RP MEDLINE=99021390; PubMed=9806553;
 RX Robertson N.G., Lu L., Heller S., Merchant S.N., Bayev R.D.,
 RA McKenna M., Nadol J.B. Jr., Miyamoto R.T., Linthicum F.H. Jr.,
 RA Neto J.F.L., Hudspeth A.J., Seidman C.E., Morton C.C., Seidman J.G.;
 RA "Mutations in a novel cochlear gene cause DFNA9, a human nonsyndromic
 deafness with vestibular dysfunction.",
 RA Nat. Genet. 20:299-303(1998).
 RL [3]
 RN VARIANT DFNA9 SER-51.
 RP MEDLINE=99135917; PubMed=991344;
 RX de Kok Y.J.M., Bom S.J.H., Brunt T.M., Kemperman M.H.,
 RA van Beusekom E., van der Velde-Visser S.D., Robertson N.G.,
 RA Morton C.C., Huygen P.L.M., Verhagen W.I.M., Brunner H.G.,
 RA Cremers C.W.R.J., Cremers F.P.M.;
 RA "A proxi-to-ser mutation in the COCH gene is associated with late
 onset autosomal dominant progressive sensorineural hearing loss with
 vestibular defects.",
 RA Hum. Mol. Genet. 8:361-366(1999).
 RL [4]
 RN VARIANT DFNA9 ASN-109.
 RP Kamariotis M., McGill J., Lynch M., Dahl H.-H.M.;
 RA "Genetic studies of a family with inherited dominant hearing loss.",
 RA Unpublished observations (JAN-2000).
 RL [5]
 CC -1- SUBCELLULAR LOCATION: secreted (potential).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN INNER EAR STRUCTURES; THE COCHLEA
 CC AND THE VESTIBULE.
 CC -1- DISEASE: DEFECTS IN COCH ARE THE CAUSE OF AUTOSOMAL DOMINANT
 CC NONSYNDROMIC SENSORINEURAL DEAFNESS TYPE 9 (DFNA9). AFFECTED
 CC INDIVIDUALS HAVE MICROPOLYSACCHARIDE DEPOSITIONS IN THE CHANNELS OF
 CC THE COCHLEAR AND VESTIBULAR NERVES. THESE DEPOSITIONS APPARENTLY
 CC CAUSE STRANGULATION AND DEGENERATION OF DENDRITIC FIBERS.
 CC -1- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
 CC -1- DATABASE: NAME=Protein Spotlight;
 CC NOTE=Issue 4 of November 2000;
 CC WWW="http://www.expasy.org/spotlight/articles/spilt004.html";
 CC -1- DATABASE: NAME=Hereditary hearing loss homepage;
 CC WWW="http://www.uia.ac.be/dnalab/hhh/hhgenes.html";
 CC NOTE=Gene page;
 CC WWW="http://www.uia.ac.be/dnalab/hhh/hhgenes.html";
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL; AF006740; AAC39545.1; -
 DR HSSP; P11215; LJM.
 DR Genew; HGNC:2180; COCH.
 DR MIM; 603196; -
 DR MIM; 601369; -
 DR InterPro; IPR004043; LCCL_dom.
 DR InterPro; IPR02035; VWFA.
 DR Pfam; PF00092; vwa; 2.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00327; VWA; 2.
 DR PROSITE; PS50820; LCCL; 1.
 DR PROSITE; PS50234; VWFA; 2.
 KW Glycoprotein; Repeat; Signal; Disease mutation; Deafness;

CC	IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC	-1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-slb.ch/announce/ or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL; D50086; BAA08789.1; -
DR	HSSP; P12259; IC2T.
DR	MGI; MG1:106206; NTP.
DR	InterPro; IPR000859; CUB_domain.
DR	InterPro; IPR000421; FA58_C.
DR	InterPro; IPR000998; MAM_domain.
DR	Pfam; PF00431; CUB; 2.
DR	Pfam; PF00629; MAM; 1.
DR	Pfam; PF00754; F5_F8_type_C; 2.
DR	SMART; SM00042; CUB; 2.
DR	SMART; SM00231; FA58C; 2.
DR	SMART; SM00137; MAM; 1.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS01285; FA58C_1; 2.
DR	PROSITE; PS01286; FA58C_2; 2.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PS50060; MAM_2; 1.
KW	Transmembrane; Glycoprotein;
FT	SIGNAL 1 21
FT	CHARIN 22 923
FT	DOMAIN 22 856
FT	TRANSERM 857 879
FT	DOMAIN 880 923
FT	DOMAIN 27 141
FT	DOMAIN 147 265
FT	DOMAIN 275 424
FT	DOMAIN 431 583
FT	DOMAIN 645 811
FT	DISULFID 27 34
FT	DISULFID 82 104
FT	DISULFID 147 173
FT	DISULFID 206 228
FT	DISULFID 275 424
FT	DISULFID 431 583
FT	CARBOHYD 150 150
FT	CARBOHYD 261 261
FT	CARBOHYD 300 300
FT	CARBOHYD 522 522
FT	CARBOHYD 842 842
SO	SEQUENCE 923 AA; 103020 MW; 0644B8A170796808 CRC64;
Query Match	5.48; Score 143.5; DB 1; Length 923;
Best Local Similarity	19.38; Pred. No. 0.0033;
Matches 104; Conservative	80; Mismatches 199; Indels 157; Gaps 24;
QY	21 LRLAELGDCGHIIVTSODSGTMSKNKYPCGYPRVYCEKIIIVYPKKRLILRGDINI 80
DB	136 IRYELFKRGPCSONYTR-PTGVYIKSPFEPKRYNCLCTYIIIPAKKSEIILEESFDL 194
QY	81 ESKT-----CASDYLLFSSATDQYD---YCGSWAVPEKLRLNSNEYVLFKSGSHI 129
DB	195 EODSNPGCMGCRFYDRLEIMDGFPEVGHIGRYCGQ-KTPGRIRSSGVLISWVFYDGA 253
QY	130 SGREFLLTYA---SSDRHDLITCLEGRSHYFEKYSKCPAGCARDIADISGNTKDGIRD 186
DB	254 AKKEGFSANYSVYLGSSISSEDF-----KCEMALGMEGSE----- 285

QY 187 TSLCKAAIHAGITT--DELGHIINLQSK-----GISHYE-----GLT-----223
 DB 286 -----IHSDQITASSQGTNMVSERSLANTPENCGWTGEGDEYKEMIOVDGLLFVET 337
 QY 224 ---ANGVLSHSGLSERFLFTPGANITVALPSYF-----TALLITG-----MGI 268
 DB 338 AVGTGQAISE---TKKKYVKYRVDISNGEDWLSLKGKNAIIFGQWTFPTDVLGV 394
 QY 269 FA-----ICRRKKKGNPNVSADAOKTGCWKQIKYFPARH-----OSTEFTISYDN 314
 DB 395 FSRPLTRFPRIRPVSMWEGISMRFEYGC-KITDIPCSGMLGWSGLISDSQITASSNOA 453
 QY 315 EKE-WTQKIDLITS-----DNADYO--OPLMIGTG-----341
 DB 454 DRMMMPENIRLTSRTGMALEPSPHPYTNEMLOVDLGEDEKIVGVIIQGGKHENKVEYMR 513
 QY 342 ---TAAKRGSTRPMDTDEEVRVNTESAGHDCRRPRRHEVALPLTISEEYATPT- 336
 DB 514 KFRIVASNNQSDMKTTIMDSKRRAKSEEGNNNDTPELRTFSPISFRIRIYERATVHSG 573
 QY 397 --VERHLRAHFTFSTQSGYVPGPRPTHK-----HSHSGGFPATGATGVESYORP 446
 DB 574 LGIRMELGCEVAPTAGPTTPNGNPNVHECDDOANCHSGTGDF-QLTGCTVLAETKP 632

RESULT 10
 BML1_MOUSE
 ID BML1_MOUSE STANDARD: PRT: 991 AA.
 AC P98063:
 DT 01-FEB-1996 (rel. 33, Created)
 DT 01-FEB-1996 (rel. 33, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Bone morphogenetic protein 1 precursor (EC 3.4.24.19) (BMP-1)
 GN (Procollagen C-proteinase) (PCP) (Mammalian tolloid protein) (mtld).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RZ MEDLINE=94229342; PubMed=8174772;
 RA Fukagawa M., Noboru S., Hogan B.L.M., Jones C.M.;
 RT "Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),
 RT which is related to the Drosophila dorsventral gene tolloid and
 RT encodes a putative astacin metalloendopeptidase.";
 RL Dev. Biol. 163:175-183(1994).
 CC -1- FUNCTION: CLEAVES THE C-TERMINAL PROPEPTIDES OF PROCOLLAGEN I, II
 CC AND II. INDUCES CARILAGE AND BONE FORMATION.
 CC -1- CATALYTIC ACTIVITY: Cleavage of the C-terminal propeptide at
 CC Ala-|-Asp in type I and II procollagens and at Arg-|-Asp in type
 CC III.
 CC -1- CORFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- ENZYME REGULATION: ACTIVITY IS INCREASED BY THE PROCOLLAGEN C-
 CC ENDOPEPTIDASE ENHANCER PROTEIN.
 CC -1- TISSUE SPECIFICITY: AT HIGH LEVELS IN EMBRYONIC MATERNAL DECIDUUM
 CC AND FLOOR PLATE REGION OF THE NEURAL TUBE. LESS IN DEVELOPING
 CC MEMBRANOUS AND ENDOCHONDRAL BONE, SUBMUCOSA OF INTESTINE, DERMIS
 CC OF SKIN AND THE MESENCHYME OF SPLEEN AND LUNG.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
 CC -1- SIMILARITY: CONTAINS 5 CUB DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L24755; AAA37306.1; -

DR HSSP; P00736; IAPQ.
 DR MEROPS; M12.005; -.
 DR MGD; MGI:88176; Bmp1.
 DR InterPro: IPR001506; Astacin.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR000130; Zn_MTPeptide.
 DR Pfam; P000008; EGF; 2.
 DR Pfam; PFO0431; CUB; 5.
 DR Pfam; PFO1400; Astacin; 1.
 DR PRINTS; PR00480; ASTACIN.
 DR SMART; SM00042; CUB; 5.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00235; ZINC; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00180; CUB; 5.
 DR PROSITE; PS00010; ASX_HYDROXYL; 2.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR Growth factor; Cytokine; Repeat; Bone; Cartilage; Hydrilase;
 KW Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;
 KW Glycoprotein; Zymogen.
 FT SIGNAL; 1 25
 FT PROPEP; 26 125
 FT CHAIN; 126 991
 FT DOMAIN; 126 326
 FT DOMAIN; 327 439
 FT DOMAIN; 440 551
 FT DOMAIN; 552 593
 FT DOMAIN; 595 707
 FT DOMAIN; 708 748
 FT DOMAIN; 752 864
 FT DOMAIN; 865 981
 FT ACT_SITE; 219 219
 FT METAL; 222 222
 FT METAL; 228 228
 FT DISULFID; 327 333
 FT DISULFID; 380 402
 FT DISULFID; 440 466
 FT DISULFID; 493 515
 FT DISULFID; 556 568
 FT DISULFID; 564 577
 FT DISULFID; 579 592
 FT DISULFID; 596 622
 FT DISULFID; 649 671
 FT DISULFID; 712 723
 FT DISULFID; 719 732
 FT DISULFID; 734 747
 FT CARBOHYD; 96 96
 FT CARBOHYD; 147 147
 FT CARBOHYD; 337 337
 FT CARBOHYD; 368 368
 FT CARBOHYD; 604 604
 SQ SEQUENCE 991 AA; 111607 MW; 68A1847783A0BB9E CRC64;

Query Match 5.3%; Score 142.5; DB 1; Length 991;
 Best Local Similarity 33.0%; Pred. No. 0.0044;
 Matches 38; Conservative 22; Mismatches 48; Indels 7; Gaps 4;

QY 31 GCCHYTSDDSGMTSKNPNYGTVPNTVCKITVKGKRLILRLDGLNTESK-TASD 89
 DB 751 GCEHKYTS-TSGHITSPNWDKYPKSKKCTWALSTPGHVRVKLTPEMDIESOPBCAYDH 809
 QY 90 LFESSATDQGP-----YCGSWAVPKELRLNSNVTVLFGSGSHISGRGFLTYAS 140
 DB 810 LEVFDGRDAKAPVLRGFCGS-KRPEPVLATGNMFLRFTSDNSVQKKGFQASHST 863

RESULT 11

FT VARSPLIC 589 622 AACGGELTKINGSITSPGFKPPKNCITWOLV -> GCY
 FT DLAOVGRPLMDRHCFLSTGPEMLGTALNG (IN
 FT ISOFORM BMP1-5).
 FT VARSPLIC 623 986 MISSING (IN ISOFORM BMP1-5).
 FT VARSPLIC 703 717 DKDECKDNCGCCOD -> GGELFGILGHPRRR (IN
 FT ISOFORM BMP1-6).
 FT VARSPLIC 718 986 MISSING (IN ISOFORM BMP1-6).
 FT VARSPLIC 703 823 DKDECKDNCGCCODVNTGSEYECQRCGFVLHKNKDK
 FT EACCDHVKVSTSGITISPMMPDYKPKCECTMAISYPIGRH
 FT VKLTFEMDIESQPECAVDHLEFVDGDAKAPVLYGREG
 FT -> VLKAGDRHSHLSGLLELLPHALVDVPAAPSLHSD
 FT THAHTHYHTHPIAOTPCRGPPPLGASRLSPGPHLITAD
 FT POGSTLDFWDTIRGDPKPRKRKSLTSTPATFERGIVA
 FT L (IN ISOFORM BMP1-7).
 FT VARSPLIC 824 986 MISSING (IN ISOFORM BMP1-7).
 FT CONFLICT 748 748 D -> N (IN REF. 4).
 FT CONFLICT 934 934 R -> S (IN REF. 4).
 SQ SEQUENCE 986 AA: 111248 MW: F89201913AC3CBEA CRC64;

Query Match 5.3%; Score 141.5; DB 1; Length 986;
 Best Local Similarity 33.0%; Pred. No. 0.0052;
 Matches 38; Conservative 22; Mismatches 48; Indels 7; Gaps 4;
 QY 31 GCGHIVTSQDSGTMSTKPYPTVTCERITVPKGRILRLGDLNIEBK-TQASDY 89
 DB 746 GCHKRTS-TSGITISPMMPDYKPKCECTMAISYPIGRHKLTFEMDIESQPECAVDH 804
 QY 90 LFSNATDQGP---YCGSMVAPKELRLNSNEVTVLFKSGSHSGFLLTVAS 140
 DB 805 LEVFDGDAKAPVLYGREGS-KRPEVPLATGSRKFLRFYSDNSYORGFQASHAT 858

RESULT 12
 UVS2.XENLA STANDARD; PRT; 514 AA.
 AC P42664;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Embryonic protein UVS.2 precursor (EC 3.4.24.-).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
 OC Xenopodidae; Xenopus.
 NC NCB1_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97228908; PubMed=9074934;
 RA Katagiri C., Maeda R., Yamashita C., Mita K., Sargent T.D.,
 RA Yasunasu S.;
 RT "Molecular cloning of Xenopus hatching enzyme and its specific
 RT expression in hatching gland cells.";
 RL Int. J. Dev. Biol. 41:19-25(1997).
 RN [2]
 RP SEQUENCE OF 196-514 FROM N.A.
 RX MEDLINE=90108413; PubMed=1688538;
 RA Sato S.M., Sargent T.D.;
 RT "Molecular approach to dorsoanterior development in Xenopus laevis.";
 RL Dev. Biol. 137:135-141(1990).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- DEVELOPMENTAL STAGE: EXCLUSIVELY IN THE ANTERIOR NEURAL FOLD OF
 CC NEURULA STAGE EMBRYOS. BY THE TAILBUD STAGE, THE PROTEIN IS
 CC LOCALIZED IN SPECIALIZED CEPHALIC ECTODERM, IN A REGION PROBABLY
 CC CORRESPONDING TO THE HATCHING GLAND.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL: D89632; BA14003.1; -
 CC DR EMBL: M27162; AAA49980.1; -
 CC DR HSSP: P07584; 11AE.
 CC DR MEROPS; M12.014; -
 CC DR InterPro: IPR001506; Astacin.
 CC DR InterPro: IPR000859; CUB_domain.
 CC DR InterPro: IPR000130; Zn_MTPeptidse.
 CC DR Pfam: PF00431; CUB; 2.
 CC DR Pfam: PF01400; Astacin; 1.
 CC DR PRINTS; PR00480; ASTACIN.
 CC DR SMART; SM00235; ZNMC; 1.
 CC DR SMART; PS00142; ZINC_PROTEASE; 1.
 CC DR PROSITE; PS00180; CUB; 2.
 CC DR PROSITE; PS01180; CUB; 2.
 CC KW Hydroxylase; Metalloprotease; Zinc; Repeat; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 514 EMBRYONIC PROTEIN UVS.2.
 FT DOMAIN 288 400 CUB 1.
 FT DOMAIN 402 513 CUB 2.
 FT METAL 186 186 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 187 187 BY SIMILARITY.
 FT METAL 190 190 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 196 196 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 514 AA: 56844 MW: DE1FCB599014D986 CRC64;

Query Match 5.3%; Score 141; DB 1; Length 514;
 Best Local Similarity 34.5%; Pred. No. 0.0023;
 Matches 38; Conservative 11; Mismatches 57; Indels 4; Gaps 3;

QY 32 GCHIVTSQDSGTMSTKPYPTVTCERITVPKGRILRLGDLNIE-SKTCASDYL 90
 DB 402 CGCAFYSPPK-TFSPYPCNVYTTNCTWTTPAPAGFKVSLRTDELEIGASCARDYL 460
 QY 91 LFSNAT--DOYPCGSMVAPKELRLNSNEVTVLFKSGSHSGFLLTV 138
 DB 461 NIVSTLGAWPGYCPIDFHSALVSKNSMTMNSDFSKOYKGFSAATY 510

RESULT 13
 BMPH_STRPU STANDARD; PRT; 639 AA.
 AC P98069;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bone morphogenetic protein 1 homolog precursor (EC 3.4.24.-) (SUBMP).
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 NC NCB1_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94215496; PubMed=8162855;
 RA Hwang S.P.L., Parin J.S., Lennarz W.J.;
 RT "Characterization of a homolog of human bone morphogenetic protein 1
 RT in the embryo of the sea urchin, Strongylocentrotus purpuratus.";
 RL Development 120:559-568(1994).
 CC -1- TISSUE SPECIFICITY: ECTODERMAL AND PRIMARY MESENCHYM CELLS IN
 CC HATCHED BLASTULA.
 CC -1- DEVELOPMENTAL STAGE: EMBRYO, HIGHEST LEVEL BEFORE SPICULOGENESIS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; L23838; AAA30081.1; -
 DR HSSP; P00736; IAPQ.
 DR MEROPS; M12.005; -
 DR InterPro; IPR001506; Astacin.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_ca.
 DR InterPro; IPR000130; Zn_MTpeptidase.
 DR Pfam; PF000431; CUB; 2.
 DR Pfam; PF01400; Astacin; 1.
 DR PRINTS; PR00480; ASTACIN.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00235; ZNMG; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 KW Developmental protein; Repeat; Hydrolase; Protease; Zinc; Embryo;
 KW Metalloprotease; EGF-like domain; Calcium; Signal.
 FT SIGNAL 1 23
 FT PROPEP 1 109
 FT CHAIN 110 639 BONE MORPHOGENETIC PROTEIN 1 HOMOLOG.
 FT DOMAIN 110 306 METALLOPROTEASE.
 FT DOMAIN 307 419
 FT DOMAIN 420 531 CUB 1.
 FT DOMAIN 532 573 CUB 2.
 FT METAL 197 197 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 FT ACT_SITE 198 198 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 201 201 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 207 207 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 536 548
 FT DISULFID 544 557 BY SIMILARITY.
 FT DISULFID 559 572 BY SIMILARITY.
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 639 AA; 71893 MW; 59307B265B7894AD CRC64;

Query Match 5.3%; Score 141; DB 1; Length 639;
 Best Local Similarity 26.2%; Pred. No. 0.0031;
 Matches 50; Conservative 28; Mismatches 63; Indels 50; Gaps 10;

33 CHVTSDDSGMTSKNTPTGTYPTVCEKIIIVPKGRLLRLADLNTIE-SKTCASDYLL 91
 422 GHI--ERSSGTLQSPNYPDDHPKSECVMLTTPANLYVVGSGFSEFIERHETCIYDYVE 479
 92 FSSATDQGP---YCGSWAYPKELRLNSNEVTYLFKSGHSISRGFLIYTAASGHPDLI 147
 480 VRDGHEDTSPILGRYCG-YFIPDDIKSTGNMAMVTFVSDGSVNRGGF-----SAD----- 528
 148 TGLRGSHYFEKYSKCP--AGCRDIAGDISGN-----TKDGYRDTSLCKNA 194
 529 -----FFEKDECAQPDGGGCMQVNTIGSYRDCRCRYELSDGRR-----CEVA 575
 195 -----IHAGT 199
 576 AEVYSLVEGI 586

RESULT 14
 ID PCOL_RAT STANDARD; PRT; 468 AA.
 AC 008628;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Procollagen C-proteinase enhancer protein precursor (PCPE) (Type I
 DE procollagen COOH-terminal proteinase enhancer) (Type I procollagen C-
 DE proteinase enhancer protein).
 GN PCOLCE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar;
 RX MEDLINE=97447750; PubMed=9303490;
 RA Ogata I., Auster A.S., Matsui A., Greenwel P., Geerts A., D'Amico T.,
 RA Fujiwara K., Kessler E., Rojkind M.;
 RT "Up-regulation of type I procollagen C-proteinase enhancer protein
 RT messenger RNA in rats with CCl4-induced liver fibrosis.";
 RL Hepatology 26:611-617(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Aorta;
 RA Hirahara I., Syoufuda K., Harada K., Tomita M., Drakami K., Terai H.,
 RA Mitsuaki N., Saito Y.;
 RT "Smooth muscle cell derived procollagen C-protease enhancer protein.";
 RL Cell Struct. Funct. 21:662-662(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer 344;
 RA Masuda M., Igarashi H., Kano M., Yoshikura H.;
 RT "Proximal integration into the procollagen C-proteinase enhancer
 RT protein gene and its effects in cultured rat fibroblasts revealed by
 RT an excisable 'hit-and-run' retroviral vector.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=91076903; PubMed=2256940;
 RA Kessler E., Mould A.P., Holmes D.J.S.;
 RT "Procollagen type I C-proteinase enhancer is a naturally occurring
 RT connective tissue glycoprotein.";
 RL Biochem. Biophys. Res. Commun. 173:81-86(1990).
 CC -1- FUNCTION: BINDS TO THE COOH-TERMINAL PROPEPTIDE OF TYPE I
 CC PROCOLLAGEN AND ENHANCES PROCOLLAGEN C-PROTEINASE ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN COLLAGEN-RICH
 CC TISSUES, ESPECIALLY TENDON. ALSO EXPRESSED IN CORNEA AND STERNA.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 NTR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

FT DISULFID 765 778 BY SIMILARITY.
FT DISULFID 783 809 BY SIMILARITY.
FT DISULFID 836 858 BY SIMILARITY.
FT DISULFID 896 926 BY SIMILARITY.
FT DISULFID 953 975 BY SIMILARITY.
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1022 AA; 115536 MW; A68CALDDE41793F9 CRC64;

Query Match 5.28; Score 139; DB 1; Length 1022;
Best Local Similarity 29.08; Pred. No. 0.0086;

Matches 40; Conservative 23; Mismatches 45; Indels 30; Gaps 7;

QY 32 CGHIVTSODS-GMTWSKNYPGYPNYTVCEKITVPRKRLILRLGLDNI-ESKTCASDY 89
DB 358 CGB--TLQDSVGNFSSPGIPNGIPSYTHCWMRISVTPGKIVLNTFTMDLYKSSLCWYDY 415
QY 90 LFFSSATDQY-----GPGCGMAVPKELRLNSNEVTVLFKSGSHISGRFLTY---- 138
DB 416 I---EVRDGYWRKAPLGRFCGD-KIPEYLVSTDSNMWTEFRSSSNWVGCPAAYEAIC 471
QY 139 -----ASSDHPD 145
DB 472 GGEISKDSGOIQSPNYPD 489

Search completed: May 15, 2003, 13:16:10
Job time : 17.6281 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:37 ; Search time 49.1437 Seconds
(without alignments)
2108.953 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGSPVIALFLVCAAP.....YSAPRNGIAPLNOTAMTALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPRMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriophage.*
17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2665	99.8	503	11	Q9D4J3 mus musculus
2	2448	91.7	460	11	Q9D696 mus musculus
3	2239.5	83.8	432	11	Q8R327 mus musculus
4	575.5	21.5	775	4	Q96PD2 mus musculus
5	575.5	21.5	775	4	Q8TDX2 mus musculus
6	565.5	21.2	769	11	Q91ZV2 mus musculus
7	560	21.0	769	11	Q91ZV3 mus musculus
8	503	18.8	155	4	Q96NH2 mus musculus
9	418	15.6	364	4	Q14089 mus musculus
10	178	6.7	414	11	Q9CX06 mus musculus
11	178	6.7	414	11	Q8R4W6 mus musculus
12	172	6.4	414	11	Q8BY4 mus musculus
13	170.5	6.4	415	4	Q9UKZ9 mus musculus
14	169.5	6.3	694	11	Q70542 mus musculus
15	169.5	6.3	694	11	Q8R099 mus musculus
16	168	6.3	1524	13	Q91674 xenopus lae

17	167.5	6.3	3623	11	Q70244 mus musculus
18	165	6.2	541	11	Q9GX90 mus musculus
19	165	6.2	623	11	Q9JUP3 mus musculus
20	165	6.2	643	11	Q9GX84 mus musculus
21	165	6.2	678	11	Q9JUS8 mus musculus
22	162.5	6.1	3623	4	Q60494 mus musculus
23	162	6.1	685	11	Q9Z338 mus musculus
24	160	6.0	685	11	Q91WP0 mus musculus
25	159.5	6.0	701	11	Q9JUS9 mus musculus
26	159.5	6.0	733	11	Q92050 mus musculus
27	159.5	6.0	746	5	Q01654 mus musculus
28	157.5	5.9	3620	6	Q9T053 mus musculus
29	157	5.9	276	4	Q9BRH3 mus musculus
30	157	5.9	384	13	Q9W630 mus musculus
31	154	5.8	251	11	Q9D9K5 mus musculus
32	154	5.8	745	13	Q9PVY3 mus musculus
33	153.5	5.7	728	4	Q96R54 mus musculus
34	152.5	5.7	719	13	Q9PVY2 mus musculus
35	151.5	5.7	919	13	Q8UVR0 mus musculus
36	151.5	5.7	936	13	Q8UVQ9 mus musculus
37	150	5.6	1594	6	Q95218 mus musculus
38	148	5.5	926	4	Q9UQ00 mus musculus
39	147	5.5	1015	4	Q9Y6L7 mus musculus
40	145	5.4	494	4	Q961U6 mus musculus
41	145	5.4	693	4	Q96DM8 mus musculus
42	144.5	5.4	722	13	Q9PSZ5 mus musculus
43	144	5.4	650	11	Q8VH15 mus musculus
44	144	5.4	977	13	Q91925 mus musculus
45	143	5.4	579	4	Q96DQ9 mus musculus

ALIGNMENTS

RESULT 1	ID	Q9D4J3	PRELIMINARY;	PRT;	503 AA.
AC	Q9D4J3	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	4631413K1R1K protein.				
GN	4631413K1R1K.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;				
RX	MEDLINE=5781/6J; PubMed=11217851;				
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,				
RA	Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Koehle H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,				
RA	Hashizaki Y.,				
RT	*Functional annotation of a full-length mouse cDNA collection.*				
RL	Nature 409:685-690(2001).				
CC	-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.				
DR	EMBL; AK016485; BAB30265.1; -				

DR MGD: MGI:1913936; 4631413K1R1K.
 DR InterPro: IPR000859; CUB_dom.
 DR InterPro: IPR004043; ICCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SO SEQUENCE 503 AA; 54547 MW; F8B121E845CAA06B CRC64;

Query Match 99.8%; Score 2665; DB 11; Length 503;
 Best Local Similarity 99.8%; Pred. No. 1.5e-219;
 Matches 502; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGTGAGSPVALAFVACAPRLQAEEIGDGGHVTSDSGTMSKVPCTPYATYCE 60
 DB 1 MGTGAGSPVALAFVACAPRLQAEEIGDGGHVTSDSGTMSKVPCTPYATYCE 60
 QY 61 KITVPKGRILRLRGDNIESTKTCASDYLLFSSATDQYCGYCSWAPKELRLNSNEVT 120
 DB 61 KITVPKGRILRLRGDNIESTKTCASDYLLFSSATDQYCGYCSWAPKELRLNSNEVT 120
 QY 121 VLFSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSKFCPCACRDIAIGSNT 180
 DB 121 VLFSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSKFCPCACRDIAIGSNT 180
 QY 181 KDGRDTSILCKAIIHGIITDELGHINILQSGISHYEGILANGVLSRRGSISEKREL 240
 DB 181 KDGRDTSILCKAIIHGIITDELGHINILQSGISHYEGILANGVLSRRGSISEKREL 240
 QY 241 FTFPGMNTTVAIPSVIFIALLTGMGIFAICRRKKKGNPYVADAKTCGKQIKYKF 300
 DB 241 FTFPGMNTTVAIPSVIFIALLTGMGIFAICRRKKKGNPYVADAKTCGKQIKYKF 300
 QY 301 ARHOSTETTSYDNEKEKTKLIDTSDMADYQOPLMIGTGVARKSGTFPMOTDEEV 360
 DB 301 ARHOSTETTSYDNEKEKTKLIDTSDMADYQOPLMIGTGVARKSGTFPMOTDEEV 360
 QY 361 RVNTEASGHYCCPRPGRHEVALPLTSEPEYATPIVERHLLRAHTEFSTOGSYRPGRP 420
 DB 361 RVNTEASGHYCCPRPGRHEVALPLTSEPEYATPIVERHLLRAHTEFSTOGSYRPGRP 420
 QY 421 THKSHSSGCPPTATGATQVESYORPASPKPVGGGYDKPASSFSDNDPASQSGMTSGG 480
 DB 421 THKSHSSGCPPTATGATQVESYORPASPKPVGGGYDKPASSFSDNDPASQSGMTSGG 480
 QY 481 DDGYSAPRNGIAPLNGTAMTALL 503
 DB 481 DDGYSAPRNGIAPLNGTAMTALL 503

RESULT 2
 Q9D696 PRELIMINARY; PRT; 460 AA.
 ID Q9D696
 AC Q9D696;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 4631413K1R1K protein.
 GN 4631413K1R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SKIN;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shlnagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriber L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Bono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 FT *Functional annotation of a full-length mouse cDNA collection.*;
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK014521; BAB29409.1;
 DR MGD: MGI:1913936; 4631413K1R1K.
 DR InterPro: IPR000859; CUB domain.
 DR InterPro: IPR004043; ICCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SO SEQUENCE 460 AA; 50334 MW; 3CF3356F70A88B93 CRC64;

Query Match 91.7%; Score 2448; DB 11; Length 460;
 Best Local Similarity 100.0%; Pred. No. 5.1e-201;
 Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 MTSKNYPTGYNTYCEKITVPKGRILRLRGDNIESTKTCASDYLLFSSATDQYGYPC 103
 DB 1 MTSKNYPTGYNTYCEKITVPKGRILRLRGDNIESTKTCASDYLLFSSATDQYGYPC 60
 QY 104 GSMVAPKELRLNSNEVTLYFGSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSK 163
 DB 61 GSMVAPKELRLNSNEVTLYFGSGSHISGRGFLLYASSDHPDLITCLERGSHYFEERYSK 120
 QY 164 FCPAGCRDIADISGNTRKDYRDTSLCKAIIHGIITDELGHINILQSGISHYEGIL 223
 DB 121 FCPAGCRDIADISGNTRKDYRDTSLCKAIIHGIITDELGHINILQSGISHYEGIL 180
 QY 224 ANGVLSRRGSISEKRELFTTTPGMNTTVAIPSVIFIALLTGMGIFAICRRKKKGNPY 283
 DB 181 ANGVLSRRGSISEKRELFTTTPGMNTTVAIPSVIFIALLTGMGIFAICRRKKKGNPY 240
 QY 284 SADAKTCGKQIKYPPARHOSTETTSYDNEKEKTKLIDTSDMADYQOPLMIGTGV 343
 DB 241 SADAKTCGKQIKYPPARHOSTETTSYDNEKEKTKLIDTSDMADYQOPLMIGTGV 300
 QY 344 ARKSTFPMOTDEEVYRVNTEASGHYDCPRPGRHEVALPLTSEPEYATPIVERHLLR 403
 DB 301 ARKSTFPMOTDEEVYRVNTEASGHYDCPRPGRHEVALPLTSEPEYATPIVERHLLR 360
 QY 404 AHTEFSTOGSYRVPGRPTTHKSHSSGCPPTATGATQVESYORPASPKPVGGGYDKPAASS 463
 DB 361 AHTEFSTOGSYRVPGRPTTHKSHSSGCPPTATGATQVESYORPASPKPVGGGYDKPAASS 420
 QY 464 FLDSRDPASQSGMTSGGDDGYSAPRNGIAPLNGTAMTALL 503
 DB 421 FLDSRDPASQSGMTSGGDDGYSAPRNGIAPLNGTAMTALL 460

RESULT 3
 O8R327 PRELIMINARY; PRT; 432 AA.
 ID O8R327
 AC O8R327;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similar to Riken cDNA 4631413K11 gene.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC026771; AAH26771.1;
 SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;

Query Match 83.8%; Score 2239.5; DB 11; Length 432;
 Best Local Similarity 85.5%; Pred. No. 3.3e-183;
 Matches 430; Conservative 1; Mismatches 1; Indels 71; Gaps 1;

QY 1 MGTGAGPSYALILFAVCAPRLQAEELGDCGHIYVSODSGITMSKPYETNYVCE 60
 Db 1 MGTGAGPSYALILFAVCAPRLQAEEL----- 28
 QY 61 KIIVPAGKRLILRLGDLNIESKTCASDYLLFSATDQYGPYCGSMAVPEKRLNSMEY 120
 Db 29 -----GPGCSMAVPEKRLNSMEY 49
 QY 121 VLEKSGSHISGRGFLTYASDHPDLITCLERGSYFEERYSKFCPACGDIADISGNT 180
 Db 50 VLEKSGSHISGRGFLTYASDHPDLITCLERGSYFEERYSKFCPACGDIADISGNT 109
 QY 181 KDGYRDSILCKAIIHAGIITDELGSHINILQSKGISHYEGLLANGVLSRHSLSERKFL 240
 Db 110 KDGYRDSILCKAIIHAGIITDELGSHINILQSKGISHYEGLLANGVLSRHSLSERKFL 169
 QY 241 FTTPGMMITVAIPSVIFIALLLTGMCIFAICRRKKRKNPYVSADAKTGCKMCKIYYPF 300
 Db 170 FTTPGMMITVAIPSVIFIALLLTGMCIFAICRRKKRKNPYVSADAKTGCKMCKIYYPF 229
 QY 301 ARHOSFTTISYNEKEMTKLDLITSDMADYQOPLMIGTGVARKSGTFRPMDTDEEV 360
 Db 230 ARHOSFTTISYNEKEMTKLDLITSDMADYQOPLMIGTGVARKSGTFRPMDTDEEV 289
 QY 361 RVNTEASGHDCPRPGRHEVALPLTSEPEVATPIYERHLRLHNTSGTSGYRVGPRP 420
 Db 290 RVNTEASGHDCPRPGRHEVALPLTSEPEVATPIYERHLRLHNTSGTSGYRVGPRP 349
 QY 421 TTKHSHSGGFPATGATQVESYORPASPRVGGYDPRPAASSFLDSRDPASQOMTSGG 480
 Db 350 TTKHSHSGGFPATGATQVESYORPASPRVGGYDPRPAASSFLDSRDPASQOMTSGG 409
 QY 481 DDGYSAPRNGIAPLNGATMALL 503
 Db 410 DDGYSAPRNGIAPLNGATMALL 432

RESULT 4
 Q96PD2 PRELIMINARY; PRT; 775 AA.
 AC Q96PD2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 GN ESDN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX PubMed-11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT *ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is Up-regulated after Vascular Injury.*
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1. SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 DR EMBL: AF387547; AAL30178.1; -

DR Interpro; IPR000859; CUB_domain.
 DR Interpro; IPR000421; FA58_C.
 DR Interpro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_F8_Type_C; 1.
 DR PROSITE; PS01180; CUB; 1.
 DR DNA-binding; Transcription regulation.
 SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Query Match 21.5%; Score 575.5; DB 4; Length 775;
 Best Local Similarity 24.5%; Pred. No. 2.4e-40;
 Matches 183; Conservative 84; Mismatches 181; Indels 299; Gaps 20;

QY 8 PSYALILFAVCAPRLQAEELGDCGHIYVSODSGITMSKPYETNYVCEIIVPK 67
 Db 48 PLFLILLVLLLEDDAGAOQDCCGHTVLAGPESGITLSTNYPTYNSTVCEIIVK 107
 QY 68 GKRLILRLGDLNIE-SFTCASDYLLFSS---ATDQYPCG-SMAVPEKRLNSMEY 121
 Db 108 GERVRIKFGDFDIBSDSCHFNLYRIYNGIGVSRTKGYGGLQNMHSIESKNEITL 167
 QY 122 LFKSGSHISGRGFLTYASDHPDLITCLERGSYFEERYSKFCPACGDIADISGNT 181
 Db 168 LFKSGSHISGRGFLTYASDHPDLITCLERGSYFEERYSKFCPACGDIADISGNT 227
 QY 182 DGYRDSILCKAIIHAGIITDELGSHINILQSKGISHYEGLLANGVLSRHSLSERKFL 241
 Db 228 HGYRDSILCKAIIHAGIITDELGSHINILQSKGISHYEGLLANGVLSRHSLSERKFL 287
 QY 242 TT----- 243
 Db 288 KTSQCYGLMGESVIADPOITASSVLEMTDHTGQENSMKPKARKLKPMPMAAFATDE 347
 QY 244 -----PGM----- 246
 Db 348 YQMLQIDLNKEKITGITTTGSTMVHENYYSAYRILYSDGQKWTYRBEVEQDKIFQ 407
 QY 247 ----- 246
 Db 408 GNMQYHGVANNFLPPIIARFIRNPYQWQOKIAMKMLLGCQFIPKGRPKLQPPPPR 467
 QY 247 ----- 246
 Db 468 NSNDLKNITAPPKIANKRAKFTQPLQPRSSNEPQTEQTASPDIRNTTVPNTKQV 527
 QY 252 AIPSVI--FALLLTGMCIFAIC-----RKRKKGN--PYVSADAKTGCKMCKIYYPF 300
 Db 528 ALAAVLVPLVAVLTLLILLVCAAMHNRKKKTEGYDLPY---WDRAGMKGKMQFL 563
 QY 301 -----ARHOSFTTISYNEKEMTKLDLITSDMADYQOPLMIGTGVARKSGT 349
 Db 584 PAKAVDHEETPVARRSSE--VNHSRPEVT--TVLQADAEVAQPLVGVIGTGHIRST 638
 QY 350 FRPMDTDEEVRYNTEASG-----YDCPRPGRHEVALPLTSEPEVATPIYERHLRLA 404
 Db 639 FRPMDTDEEVRYNTEASG-----YDCPRPGRHEVALPLTSEPEVATPIYERHLRLA 404
 QY 405 HTFSTGSGYRVPG-PRPTHSHSGGFPATGATQVESYORPASPRVGGYDPRPAAS 462
 Db 683 ---MDSGHPTTSYVGQSTSTFKATGNQPPPLVGTNTLLSRDSCSSAOAYDTFPAK 739
 QY 463 SFLDSRD-----PASQOMTSGGDDG 483
 Db 740 SFLDSRD-----PASQOMTSGGDDG 766

RESULT 5
 Q8TDX2 PRELIMINARY; PRT; 775 AA.
 AC Q8TDX2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

QY 254 PSYI--FIALLTGGMGIFALC-----RRKKKG-NPIVSAQAOTGCKKQIK--YPPARHQ 304
 Db 525 AALVLPVLAVALTLLILVCAWHNRKRKAEGTYDLPMDRAGMKGKOLLPAKSVE 584
 QY 305 STEFTISYDNE-----KEMTOKLIDITSDMADYOOPLMIGTGTVARKSTPRMDT 357
 Db 585 HEETPRVINSSEVSHLSPREVT---YVQADSAEYADPLVGGIVGTLHQRSTFRP----- 636
 QY 358 EEEVNVTEAS-----GHYDCPHRPGRHETALPLTHSEPEVATPIVERHILRAHFTSTOSGY 413
 Db 637 EE---GNEASYADLDIPYNAVQEVYAHAYAPLPTVGTGEVATPIV-----MDMSGH 683
 QY 414 RVVG-PRPTKHHSSSGGFPATGATQVSEYQRPASPKPVGGYDKP-----AASSFLD 466
 Db 684 STRASVGLPSTSTRTAGNOPALVGYNTLLSTRDSCSSQAOYDTPKGRKPAAPPELY 743
 QY 467 SRDPASOSQMTSGDD 482
 Db 744 YQVPOSTQEAAGAGRD 759

RESULT 7

ID 091ZV3 PRELIMINARY; PRT; 769 AA.
 AC 091ZV3;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 GN ESDN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR;
 RX PubMed=11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N., Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT *ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the Longest Secretory Signal Sequence among Eukaryotes, is up-regulated after Vascular Injury.*
 RT J. Biol. Chem. 276:34105-34114(2001).
 RL -1. SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 CC EMBL: AF387548; AL30179.1; -;
 DR MGD: MGI:1920639; ESDN.
 DR InterPro: IPR000859; CUB_domaln.
 DR InterPro: IPR000421; FASB_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 DR DNA-binding, transcription regulation.
 KW SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Query Match 21.0%; Score 560; DB 11; Length 769;
 Best Local Similarity 24.6%; Pred. No. 5e-39;
 Matches 191; Conservative 73; Mismatches 167; Indels 344; Gaps 23;

QY 4 GAGGSPV-----LALLFAVCAPL-----RL-----QAEELGDCGCHIV 37
 Db 15 GREGPAAPATGAAALPSAGCCPLPGRNSSSRLLLLLLLLLLODAGGGGDCGCHIV 74
 QY 38 SDDSGMTSKNYRGTYPNTVCEKITTVERGKLLRLGLDLNIE-SKTCAADYL-LPSS- 94
 Db 75 GPSSGLTSLINPHTYPSNVCEWEIRVTRGERIRIKRFGFDIEDSDYCHLNTKIFNGI 134
 QY 95 --ATDQGYPCG-SMAVPEKRLNSNEVYLFKSGHISGFGFLTYASDDHDLITCLE 151
 Db 135 GVSRTIIGKTCGIGLQMNOSIESKGEVYVLPFMSGTHAAGRGFLASTYVIDKEDLITCLD 194

QY 152 RGSHPFEKSKFCPCACRDIADISGNTKDGYNDSLCKAAIHAGIITDELGHINL 211
 Db 195 TVSNFLPEPESKYPACCLPFAEISTGTRPHGYDSSPLCMAGHAGVSNVLAGOISY 254
 QY 212 QSKGISHEGLANGVLSRSGLSSEKRLP----- 241
 Db 255 ISKGTPESSLANNVSTVGYSASLFFPTKSGCYGLMESGVADPQITASSALEWT 314
 QY 242 ----- 241
 Db 315 DHNGQENSWTAERAKRLKPGPPMAFATDEHQLDILNKKITGIVTGTSTMIERSY 374
 QY 242 ----- 241
 Db 375 VSAFVLYSDQGRWYTYRREPVQDKIFQGNKYHNDVANNFLPPIAFIRVNPQWQ 434
 QY 242 ----- 241
 Db 435 OKIAMVELLCOFTLGRLPKLTTPPRNGNLTNTARPKLGKRAPKFTQVLOPRSN 494
 QY 242 -----TPGMITTVAIIPSVI-----FIALLTGGMGIFALC-----RRK 276
 Db 495 ELVPQPAETTTTDPDKNTVT-PSVTQVALAVALVLAVALTLLILVCAWHNRK 553
 QY 277 KK-GNPVSAADAKTGCKMKQIKYF-----ARHSTFTISYDNEKEMTOKLIDL 324
 Db 554 KKEGAVDLPHMDRAGMKGKOLLPAKSVDHEETPVRYSTSE-VSHLSAREVT---TV 608
 QY 325 ITSDMADYOOPLMIGTGTVARKSTPRMDTDEEVAVNTASGH-----YDCPHRGRH 379
 Db 609 LOADSAEYADPLVGGIVGTLHQRSTFRP-----EE---GKEAGYADLDYNSPMQEVYH 659
 QY 380 EYALPLTHSEPEVATPIVERHILRAHFTSQSGYRVGPRPTKHHSSSGGFPA-----T 435
 Db 660 AYAEPLPYGPETATPIV-----MDMSG-----HPTASVGLSTSTFTK 698
 QY 436 GATQVESYQRPASPKPVGGYDKRPAASSFLDSQSQMTSGGDDGYSAPRNG 490
 Db 699 AGTQ-----PHLVGYTN-----TLISRTSCSSGQAOYDTPRGG 733

RESULT 8

ID 096NH2 PRELIMINARY; PRT; 155 AA.
 AC 096NH2;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
 DE CDNA FLJ30900 fis, clone FEBRA2005752.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;

RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto T., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Takuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
 RT *NEO human cDNA sequencing project.*
 RT Submitted (Oct-2001) to the EMBL/Genbank/DBU databases.
 DR EMBL: AK055462; BAB70926.1; -;
 SO SEQUENCE 155 AA; 16364 MW; 1C150371E07C2217 CRC64;

Query Match 18.8%; Score 503; DB 4; Length 155;
 Best Local Similarity 66.5%; Pred. No. 3.5e-35;
 Matches 103; Conservative 8; Mismatches 40; Indels 4; Gaps 3;

```

OY 353 MDTEEVNVTASGHDCPRGHEHYALPLHSEPEVATPIVERILLRAHFTSTSG 412
DB 1 MDTEAEAGVSTDAAGHDCFORAGRHEHYALPLAPPEPEVATPIYERHVRRAHFTSTSG 60
OY 413 YRVGPRPTHHSSSGGFPATG-ATOVESYORASPKPVGGCDKRAASEFL--DSRD 469
DB 61 YRVGPRPTHHSSSGGFPATG-ATOVESYORASPKPVGGCDKRAASEFL--DSRD 469
OY 470 PASOSQMT-SCGDDGYSAFAPRGLAPLNTAMTALL 503
DB 121 PDSQKPTHPGTSYSDYSAFAPRGLAPLNTAMTALL 155

RESULT 9
OY 014089 PRELIMINARY; PRT; 364 AA.
AC 014089;
DB 01-NOV-1996 (Tremblrel. 01, Created)
DB 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DB 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DB Hypothetical 40.0 kDa protein (Fragment).
DB Homo sapiens (Human).
DB Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
DB NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RA Shbata T.;
RT "unpublished.";
RL Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DB EMBL; D29810; BAI18909.1; -
DB InterPro; IPR000859; CUB_domain.
DB InterPro; IPR004042; FA58_C.
DB InterPro; IPR004043; LCCL_dom.
DB Pfam; PF00431; CUB; 1.
DB Pfam; PF00754; F5_P8_type_C; 1.
DB SMART; SM00042; CUB; 1.
DB PROSITE; PS01180; CUB; 1.
DB PROSITE; PS01180; CUB; 1.
DB Hypothetical protein.
DB NON_TER
FT 1
KW SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;
SQ

Query Match 15.68; Score 418; DB 4; Length 364;
Best Local Similarity 41.68; Pred. No. 2.4e-27;
Matches 96; Conservative 34; Mismatches 79; Indels 22; Gaps 6;

OY 30 DCGHIVTSQSGTMTSKNRYPTNYVCEKITVPRGKRLRLADLNTIE-SKTCASD 88
DB 1 DCGHIVTSQSGTMTSKNRYPTNYVCEKITVPRGKRLRLADLNTIE-SKTCASD 88
OY 89 YLFFSS---ATDQGPYCG-SMAVPKRLINSNEVTVLFKSGSHISGRGLTYVASSD 143
DB 61 YLRIYNGVSRTEIGKYGGLQINSHIESKNEITLLFMSGIHVSGRGLTASYVIDK 120
OY 144 PDLITCERGHSHYEERYSKFE-----CPAGCRDIAGDISGNTKQCYRPTSLCKRAI 195
DB 121 -----QRSYVLFGRHCIOQFPMNLSSVSTAPAGCILPFAELISGTLPHCYRSPSPICMAGV 173
OY 196 HAGITDELGHINILQSGISHYEGGLANGLVSRHGSLS-SEKRFPLFTTPG 245
DB 174 HAGVSTLGGQISVIVSKIGPIYESSLIANNVTSVGHLLIQVFYFTTSG 224

RESULT 10
OY 09CX06 PRELIMINARY; PRT; 414 AA.
AC 09CX06;
DB 01-JUN-2001 (Tremblrel. 17, Created)
DB 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DB 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DB 2400001018RLX protein.
DB PCOLCE2 OR 2400001018RLX.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DB EMBL; AK010249; BAB26794.1; -
DB MGD; MGI:193727; Pcolce2.
DB InterPro; IPR000859; CUB_domain.
DB InterPro; IPR001134; Netrin_C.
DB Pfam; PF00431; CUB; 2.
DB Pfam; PF01759; NTR; 1.
DB SMART; SM00042; CUB; 2.
DB PROSITE; PS01180; CUB; 2.
DB PROSITE; PS01180; CUB; 2.
DB SEQUENCE 414 AA; 45408 MW; E51A29A352A2CF1D CRC64;
SQ

Query Match 6.78; Score 178; DB 11; Length 414;
Best Local Similarity 23.48; Pred. No. 1e-06;
Matches 82; Conservative 51; Mismatches 97; Indels 120; Gaps 18;

OY 1 MGTGAGPSVALLFPAVCAPRLQAEELG-DCGHIYTSQSGTMTSKNRYPTNYVCE 59
DB 1 MGTGAGPSVALLFPAVCAPRLQAEELG-DCGHIYTSQSGTMTSKNRYPTNYVCE 59
OY 60 EKITVPRGKRLRLADLNTIE-SKTCASDYLTLSSADQY-----GPGGSAVAP 109
DB 59 TWKIVPEGRKVVVLFNFRIDLENOLCRIDYV-----DYNHANNGRIGRFGCTER-P 111
OY 110 KELRLNSNEVTVLFKSGSHISGRGLTYVASS----- 141
DB 112 GSIVASGKMKMTVMQMSDANTAGSGFMATYSAADPGKDRRCGRLEKPSGTFTPNMPD 171
OY 142 -DHPDLITC-----LERGSHYEERYSKFECPAGCRDIAGDISGNTKQCYRPTSLCKRAI 192
DB 172 RDYEVGYTCVWHIAPKQLIETLKFKEK-----DVE-----RDNY-----CRYDYV 212
OY 193 AAIHAGITDELGHINILQSGISHYEG----- 229
DB 213 AVENGGEVND-----AKRIGKYGDSPPVPIYERHVRRAHFTSTSG 262
OY 230 RHGSLSEKRFPLFTTPGMNITVAIPSVIFALLTLGGM-GIPAICRKRKK 278
DB 263 -HYKFRPKKFTTTTTPVTTLLPV-----TGLKPTVALCOQKCR 302

RESULT 11
OY 08R4W6 PRELIMINARY; PRT; 414 AA.
AC 08R4W6;
DB 01-JUN-2002 (Tremblrel. 21, Created)

```


DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Procollagen COOH-terminal proteinase enhancer protein 2.
 GN PCOLCE2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Steigltz B.M., Greenspan D.S.:
 RT Biochemical characterization and expression patterns of the
 RT Procollagen COOH-terminal proteinase enhancer protein 2.*;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF352788; AAL83947.1;
 KM Collagen.
 SQ SEQUENCE 414 AA; 45398 MW; 2A1A29A282A2CF1D CRC64;
 Query Match 6.7%; Score 178; DB 11; Length 414;
 Best local similarity 23.4%; Pred. No. 1e-06;
 Matches 82; Conservative 51; Mismatches 97; Indels 120; Gaps 18;
 QY 1 MGTGAGPSYTLALFANCAFLRQAELG-DGCHITYSDSGMTSKNPTGTPNTVC 59
 DB 1 MG-GASACIPCLLATAARARPPQTPPVFTGGILTG-ESGFGSGEPPGMPPSKSC 58
 QY 60 EKITYPKGRLLRLDLNIES-KTCASDYLLFSATPDY-----GPYCGSWAVP 109
 DB 59 TWKITPEGVVNLNPFIDLENDNLCRYDFV-----DYNGHANGQRIGRCGRTP-R 111
 QY 110 KEELNSNEVTVLFKSGHISGRGFLTVASS----- 141
 DB 112 GSIVASGNKMTYOMISDANAGSGFMATYGAAPDGGKGRYCGRLKPSGTFKTSNMPD 171
 QY 142 -DHPDLITC-----LENGSHFEERYSKFCPACGCDIAGDISGNTKQDYRTSLCK---- 192
 DB 172 RDPVGTCTWHTIAPRNQIETLKEFE-----DVE--RDNY-----CRDYV 212
 QY 193 AAIHAGITDELGSHIMLQSGISHYEG----- 229
 DB 213 AVENGGEVND-----AKRIKCYGDSPPVPIYSENNELLIOFLSDLSLTFADFIG 262
 QY 230 RRGSLSEKRFLETPPGKNITVAIPSYITALLITGM-GIFAICRRKKK 278
 DB 263 -HYKFRPKKPTTPTTPTVTTLPV-----TTGLKPTVALCOQKCR 302
 RESULT 12
 ID 08VBV4 PRELIMINARY; PRT; 694 AA.
 AC 08VBV4;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Hypothetical 77.4 kDa protein (Similar to complement component 1, s
 DE subcomponent).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC022123; AAH22123.1;
 DR EMBL: BC018319; AAH18319.1;

DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF000084; EGF; 1.
 DR Pfam: PF00089; Sush1; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00020; TRYR_SPC; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01187; EGF_CA; UNKNOWN_1.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
 KM Hypothetical protein.
 SQ SEQUENCE 694 AA; 77415 MW; 06AA125D5B8227F5 CRC64;
 Query Match 6.4%; Score 172; DB 11; Length 694;
 Best local similarity 25.5%; Pred. No. 7.1e-06;
 Matches 54; Conservative 41; Mismatches 85; Indels 32; Gaps 9;
 QY 15 FAVCAP---LRQAELGDGCGHYTSODSGTMSKRYPGTYPTCEKITVPRKRL 71
 DB 161 FCSCPPEYFLHDDMRNGVNCSDVFTALIGELSSPPYPPYPSNCEYQIOIQEGFOV 220
 QY 72 ILRL--GDLINE--SKTCASDYLLFSANDQYPCGS-WAVPKELRLNSNEVTVLFKS 125
 DB 221 VVTMQREDPVEPADSGNCPDLSLTFASKNQCGPYCGNFPGLTIRTSNLTGIYFOT 280
 QY 126 GSHISGRGFLITYASSDHPDLITCL-----ERGSHYEERYSKFCPACGCDIAG 174
 DB 281 DLMOQKGMKRLRY---HGPDISCAKRTANSWEPDKARYFEDVVKICVDGFEVEEG 336
 QY 175 DISGNT-----KDG-YRDTSLCKAIIHAGI 199
 DB 337 HVSSTYSTCQSDGQMSNGLKCQ-PVYCGI 367
 RESULT 13
 ID 09UKZ9 PRELIMINARY; PRT; 415 AA.
 AC 09UKZ9;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Procollagen C-terminal proteinase enhancer protein 2 (Procollagen
 DE C-proteinase enhancer protein 2).
 GN PCOLCE2 OR PCPE2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20334622; PubMed=10873381;
 RA Xu H., Acott T.S., Wirtz M.R.;
 RT Identification and Expression of a Novel Type I Procollagen C-
 RT Proteinase Enhancer Protein Gene from the Gliucoma Candidate Region on
 RT 3q21-q24.*;
 RL Genomics 66:264-273(2000).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=OSTEOARTHRITIC CARTILAGE;
 RX MEDLINE=21482651; PubMed=11597177;
 RA Kumar S., Connor J.R., Dodds R.A., Halsey W., Van Horn M., Mao J.,
 RA Sathe G.M., Mul P., Agarwal P., Badger A.M., Lee J.C., Gowen M.,

RA Lark M.W.:
 RT "Identification and initial characterization of 5000 expressed
 RT sequenced tags (ESTs) each from adult human normal and osteoarthritic
 RT cartilage cDNA libraries."
 RL Osteoarthritis cartilage 9:641-653(2001).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 DR EMBL; AF098269; AAF04621.1; -.
 DR EMBL; AY035400; AAK63128.1; -.
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR001134; Netrin_C.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF01759; NTR; 1.
 DR SMART; SM00042; CUB; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE; PS01180; CUB; 2.
 DR Collagen.
 KW SEQUENCE 415 AA; 45716 MW; B9A87FD9AA7A5B CMC64;
 SQ
 Query Match 6.4%; Score 170.5; DB 4; Length 415;
 Best Local Similarity 23.3%; Pred. No. 4.5e-06;
 Matches 99; Conservative 67; Mismatches 152; Indels 107; Gaps 22;
 Oy 4 GAGGCVALLFAVCAPRLQ--AEELDGGCHITVSODSGMTSKNTPGTTPNTVCEK 61
 Db 3 GANAAAPICLLAAATQLSRQSPRPVPTCGILTQ-ESGFIGSGEPGVYPPMSKCTW 61
 Oy 62 IIVPKGRRLRLGLDLNIES-KTCASDYLFFSSATDQY-----GPYCSMAVPE 111
 Db 62 KITVEGKVVVLFNFRIELESNLCRYFV-----DYRNGHANGRIGREGCTRR-PGA 114
 Oy 112 LRLNSNEVTVLFKSGSHISGRGLTVASS-----D 142
 Db 115 LVSSGNKKMVMOMISAPNAGNGFMFSAEPENERGDYCGGLIDRPSGFEKTPMNPDRD 174
 Oy 143 HPDLTTC---LENGSHYFEKYSKPCAG---CR-DIAGISGNTKNGYRSTSLCKAA 194
 Db 175 YPAGTTCWHYVAPKNOILEKFEKFDYERDNYCRDYVAVNGEVDARIGKYGDS 234
 Oy 195 IHAGITDELGHINLQSKGISHYEGLLANGVLSRHSLSKRLFTTPGNITVAIP 254
 Db 235 PPAPVLSR-----NELLIQFLSDL-SLADGFIGHY--IFRPKPLPTTQGPVTT-TFP 285
 Oy 255 SVIFALLITGM-GIFALCRKRRKK-----GNPYSAQAOKTCKMOKIYPPARHQBTEF 308
 Db 286 -----VTTGKLPYVALCOCKCRRTGLEGN-YCSSPFLVAG--TVITTTTRGSLHA 334
 Oy 309 TISYDN-----EKEWTKLDLITSDMAVOOPLMIGTGVARKSGTEFRPMDTD 356
 Db 335 TVSIINIKRGNLAIQAGKMSANLIVYCK-----QCPLL-----RGLNTYIMGOV 382
 Oy 357 TEVVR 361
 Db 383 GEDGR 387
 RESULT 14
 070542 PRELIMINARY; PRT; 694 AA.
 ID 070542;
 AC 070542;
 DT 01-AUG-1998 (Tremblurel. 07, Created)
 DT 01-AUG-1998 (Tremblurel. 07, last sequence update)
 DT 01-JUN-2002 (Tremblurel. 21, last annotation update)
 DE Serine protease.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98192519; PubMed=9524231;
 RA Sakai H., Nakashima S., Yoshimura S., Nishimura Y., Sakai N.,
 RA Nozawa Y.;

RT "Molecular cloning of a cDNA encoding a serine protease homologous to
 RT complement C1s precursor from rat C6 glial cells and its expression
 RT during glial differentiation."
 RL Gene 209:87-94(1998).
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; D88250; BAA25797.1; -.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.193; -.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00008; BGF; 1.
 DR Pfam; PF00084; Sush1; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00032; CCP; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS02440; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Calcium-binding; EGF-like domain; glycoprotein; Hydrolase; Protease;
 KW Repeat; Serine protease.
 KW SEQUENCE 694 AA; 77788 MW; CDC0ADFCEFFEC05 CMC64;
 SQ
 Query Match 6.3%; Score 169.5; DB 11; Length 694;
 Best Local Similarity 28.0%; Pred. No. 1.2e-05;
 Matches 52; Conservative 31; Mismatches 78; Indels 25; Gaps 7;
 Oy 15 FAVCAP---LRLQAEELDGGCHITVSODSGMTSKNTPGTTPNTVCEKITVYKGRRL 71
 Db 161 FCSCEPFEFLHNDMTGTCGVNSGDFVTLALIGEIASPNTPNPENSRCEYQIRLOEGFRL 220
 Oy 72 IL--RLGLDINIE---SKTCASDYLFFSSATDQYGPYCS--MAVPELRLNSNEVTVLFKS 125
 Db 221 VLTIRREDFVEPADSEGNCHDSLTFEAKNOQFGPYCGNGFPGPLTIRKQNTLDIVQIT 280
 Oy 126 GSHISGRFLTVASSDHPDLITC-----LENGSHYFEKYSKPCPACGRDIAG 174
 Db 281 DLTGONKGMKRLRY---HGDPICPKELISANSIMWPERAKYVFDVYKIVTCVDGEVEEG 336
 Oy 175 DISGNT 180
 Db 337 NV-GST 341
 RESULT 15
 08R099 PRELIMINARY; PRT; 694 AA.
 ID 08R099;
 AC 08R099;
 DT 01-JUN-2002 (Tremblurel. 21, Created)
 DT 01-JUN-2002 (Tremblurel. 21, last sequence update)
 DT 01-JUN-2002 (Tremblurel. 21, last annotation update)
 DE Similar to complement component 1, s subcomponent.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SALIVARY GLAND;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC027183; AAH27183.1; -.
SQ SEQUENCE 694 AA; 77713 MW; 455A9EE10C8CC12F CRC64;

SO SEQUENCE 694 AA; 77713 MW; 455A9EE10C8CC12F CRC64;

Query Match	6.38;	Score 169.5;	DB 11;	Length 694;
Best Local Similarity	28.0%;	Pred NO 1	2e-05;	

Matches 52; Conservative 31; Mismatches 78

Matches 52; Conservative 31; Mismatches 78; Indels 25; Gaps 7;

Qy	15	FAYCAP---	LTLOA	BEIDGCGHIVT	QSDSTJMSKRY	PCGTAPNYCEKIIIVP	PGKRL	71	
Db	161	FCSCPEYFLHDDMRTGC	YVNC	SGDVFTALIGELASPNY	PNDYFNSCETQIR	IQSGFRL	220		
Qy	72	IL--RLGD	LNIIE---	SKTCASDYLL	FSSANTDOYGPYCGS--	WAVPKELRLNS	EVTVLFKS	125	
Db	221	VLTIRREDVEDPEAD	SEGCNCHDSLTF	PAKNQGFSPYCGN	GFPGPLITIK	NSNLDLIVFQ	280		
Qy	126	GSHTSGRGFLITV	SSSDHPLIYC-----	LERGSHYE	EKYSKFC	PCPACG	CDIAG	174	
Db	281	DLTGQNKGMKRLRY--	HGDPIC	PKREISANSI	WEPEKARYVFK	DVVKIT	CVDGFE	VEVGS	336
Qy	175	DISGNT	180						
Db	337	NV-GST	341						

Db 161 FCSCPPEYFLHDMRTCGVNCSGDVFTALIGEIASPNYPNPYPENSRCEYQIRLQEGFRL 220

72 IL-RLDLINE--SKTCASDYLLFSSATDQYGPYCGS-WAVPKELRLNSNEVTVLFS 125

Db 221 VLTIRREDFVEPADSEGCNCHDSLTFAAKNQFGPYCGNGFPGLTIKTQSNLTLDIVEQT 280

QY 126 GSHISGRGFLTYASSDHPDLITC-----LERGSHYFEEKYSKFCPCAGCRDIAG 174

Db 281 DLTGQNKGWKLR-----HGDPICPKAISANSIWEPEKAKYFKDVKITCVDGFEVEG 336

1/5 DISGNT 180
QY ::|::|

DB 33 / NV-GST 34.1

Search completed: May 15, 2003, 13:18:51
Job time : 54.1437 secs

Job time : 54.1437 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2003, 13:14:36 ; Search time 11.5632 Seconds
(without alignments)
1279.897 Million cell updates/sec

Title: US-10-003-132-4

Perfect score: 2671
Sequence: 1 MGTGAGGPGSVTLALFAVCAP.....YSAPRNGIAPLQNTATALL 503

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfills1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164.5	6.2	415	4	US-09-032-523-2
2	143.5	5.4	923	3	US-08-936-135-6
3	142	5.3	986	4	US-08-872-757-4
4	140.5	5.3	788	1	US-08-572-225-1
5	139	5.2	1013	2	US-08-866-650-3
6	139	5.2	1013	4	US-09-021-287-3
7	139	5.2	1013	4	US-09-240-473-3
8	137	5.1	449	2	US-08-839-008-9
9	135	5.1	449	2	US-08-839-008-2
10	135	5.1	591	3	US-08-991-408-4
11	135	5.1	591	4	US-09-432-473-4
12	135	5.1	1013	2	US-08-866-650-5
13	135	5.1	1013	2	US-09-021-287-5
14	135	5.1	1013	3	US-08-991-408-2
15	135	5.1	1013	4	US-09-240-473-5
16	135	5.1	1013	4	US-09-432-473-2
17	133.5	5.0	468	4	US-09-032-523-8
18	132.5	5.0	909	3	US-08-936-135-18
19	132.5	5.0	926	3	US-08-936-135-20
20	131.5	4.9	922	4	US-09-116-473-4
21	131.5	4.9	925	4	US-09-116-473-2
22	130.5	4.9	1290	1	US-08-470-350B-2
23	129.5	4.8	901	3	US-08-936-135-22
24	129.5	4.8	906	3	US-08-936-135-24
25	129.5	4.8	909	3	US-08-936-135-8
26	129.5	4.8	909	3	US-08-936-135-10
27	129.5	4.8	914	3	US-08-936-135-12

28	129.5	4.8	926	3	US-08-936-135-14	Sequence 14, Appl
29	129.5	4.8	931	3	US-08-936-135-16	Sequence 16, Appl
30	126.5	4.7	855	2	US-09-027-337-2	Sequence 2, Appl
31	126.5	4.7	855	2	US-09-644-600-2	Sequence 2, Appl
32	126	4.7	401	2	US-08-839-008-5	Sequence 5, Appl
33	125.5	4.7	468	2	US-08-839-008-7	Sequence 7, Appl
34	125.5	4.7	666	4	US-09-341-587-1	Sequence 1, Appl
35	125.5	4.7	1785	4	US-09-341-587-3	Sequence 3, Appl
36	121	4.5	1019	1	US-08-296-014A-4	Sequence 4, Appl
37	121	4.5	1019	2	US-08-596-405-4	Sequence 4, Appl
38	121	4.5	1019	2	US-08-877-620-4	Sequence 4, Appl
39	121	4.5	1083	1	US-08-296-014A-2	Sequence 2, Appl
40	121	4.5	1083	2	US-08-596-405-2	Sequence 2, Appl
41	121	4.5	1083	2	US-08-877-620-2	Sequence 2, Appl
42	119	4.5	902	4	US-09-644-600-10	Sequence 10, Appl
43	118	4.4	102	4	US-09-374-135-7	Sequence 7, Appl
44	116.5	4.4	730	4	US-08-872-757-2	Sequence 2, Appl
45	116	4.3	1180	4	US-09-206-942-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1
US-09-032-523-2
; Sequence 2, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0479 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: RARRNOT02
; CLONE: 947429
; US-09-032-523-2

```

Query Match Similarity      6.2%  Score 164.5; DB 4; Length 415;
                          23.1%; Pred. No. 1.8e-08;
Matched 98; Conservative 67; Mismatches 153; Indels 107; Gaps 22.

      4 GAGGSPVALLAEVACAPRLQ--AEELSDGCGHIYVSDSGMTSKNRYGTVNPTVCER 61
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 GANMAVAPCLLLAAATQLSRQOSPERRPVTCGILLTG--ESGFIGSGFGVYPPNSKCTW 61
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      62 ITTVKGRKRLRLRLDNLIES-KTCASDYLLFSSATDQY-----GPGCGMAVPRE 111
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      62 KATVEGKRVVYLNFRFIDLESNLCRYDFV-----DYVNGHANGQRIGRFQGTFR-PCA 114
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      112 LKLNENEVYLFKSGSHISRGFLLTVASS-----D 142
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      115 LVSSGKMMVMQVIFDPAVNTAGNGFMAMFSAEENERKDDYCGGLDRPSSGFTTPMPKDD 174
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      143 HPDLITC----LERSHYFEERYSKFCPAG--CR-DIAGDISGNTKDDYRDTSLCKAA 194
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      175 YAGATVCWHIYAPRNQLELKEFEDVBRDNYCNDYAVANGGVNDARRIKGYCGS 234
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      195 IHAGITDELGSHIMLOSKGISHYEGLIANGVLSRHGSLSKRRLFTTPGNITVAIP 254
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      235 PPAPVYSER--NELLIQFLSDS-LTVADFGIHY--IFRPKRLPTTEQPVTT-TPP 285
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      255 SVTFALLITGM-GIAPICRKRKKR----GNPVYSAQAQKGCCKQIKYPPARRHQSTFE 308
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      286 -----VYTGKLPYVALCOQKCRRTGLEGN-YCSSDPIVLAG--VYTTTITRQGSLSHA 334
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      309 TISYDN-----EKEMTQRLDLITSDMADYQOPLMIGTGTVAARKSGSTFRPMQTD 356
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      335 TVSIIINIVEKGNLAIQAGKKNMSARLTVCK-----QCQLL-----RRGIANYIIMGV 382
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      357 TEVVR 361
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      383 GEDGR 387

RESULT 2
US-08-936-135-6
; Sequence 6, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 923 amino acids

```

```

: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
US-08-936-135-6

Query Match      5.4%; Score 143.5; DB 3; Length 923;
Best Local Similarity 19.3%; Pred. No. 1,1e-05;
Matches 104; Conservative 80; Mismatches 199; Indels 157; Gaps 24.

QY 21 LRLQAEELGDCGGHIVTSODSGTMTSKNTPGYNTVCEKIIIVPKGRRLILRLGLDNI 80
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 IYVEIFKRPESQNYTA-PTGVIKSPGPEPKYPCOLECTYIIIFAPKSEIILEESEFL 194
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 ESKT-----CASDYLFSSATDYG-----YCGSMVAPKRLRLNSNVTLLFSGSHI 129
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 195 EDDSNRPFGMFCRYDRLEMTDGFPEYGPPIGRYCCQ-KTPGKRSSGVLNVFTYDSKI 253
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 130 SGRGLLTVA---SSDHPDLITCLERGSHYFEKYSKFCPCACRDIADISGNTKDYRD 186
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 254 AKEGFSANYSVLQSSISEDF-----KCMALGMESE----- 285
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 TLLCLKAIIHAIT--DELGGHINLLOSK-----GISHE-----GLL----- 223
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 286 -----IHSQITPASSQYGTMSVERSLNYPENGWTPGDSYKKEWIOVDLGLRVT 337
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 ---ANGVLISRHGSLSEKRELFPTPGMNIITVAIPSYIF-----IALLTG-----MGI 268
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 338 AVGTQGAISKE---TKKRYVYTVRYVDLSSNGEDMISLEKGNKAIIPGNTNPIDVYLG 394
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 FA-----ICRRKKKGNPYVSADQKTCGKWKQIYPPARH-----QSTEFITSDN 314
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 395 FKPPLITRFVRLKPYSWETGISMRFEYVGC-KITDYPGSGMLGNVSGLSDQITASNQA 453
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 EKE-MTOKDLITS-----DMADYQ--OPLMIGT----- 341
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 454 DENMKMPENIRLIVTSRTGVALPPSPHPYTNEMIQVDLGDEKIVRGYIIQGGKREKRYEMR 513
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 ---TVARKGSTERPMADTDEEVRYNTEASGHYDCPHRPGRHEVALPLTHSEPEYATPI- 396
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 514 KKKIAYSNNGSDPMKTIIMDSKRAKASFEGNNNYDPPELTFEPLSTRIRIYPERATHSG 573
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 397 ---VEHHILRAHFFSFGSGRVPGRPPTRK-----HSHSGCGPPATGATQVSYQRP 446
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 574 IGLRPELLGCEYEATPAGPTTPNGNPHVHCDDQANCHSGTDDE-QLTGCTIVLATERP 632
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-08-872-757-4
: Sequence 4, Application US/08872757
: Patent No. 6258584
: GENERAL INFORMATION:
: APPLICANT: Prockop, Darwin J.
: APPLICANT: Hojima, Yoshio
: APPLICANT: Li, Shi-wu
: APPLICANT: Sieron, Aleksander
: TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
: TITLE OF INVENTION: PROCESSES, METHODS AND USES THEREOF
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmunds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION NUMBER:
: APPLICATION NUMBER: US/08/872.757

```

RESULT 4
 US-08-572-225-1
 Sequence 1, Application US/08572225
 Patent No. 5807981
 GENERAL INFORMATION:
 APPLICANT: Prockop, Darwin J.
 APPLICANT: Hojlma, Yoshio
 APPLICANT: Li, Shi-Mu
 APPLICANT: Steron, Alexander
 APPLICANT: Brenner, Mitch
 TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND ITS USE FOR
 TITLE OF INVENTION: DRUG DEVELOPMENT FOR THE TREATMENT OF DISEASE
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/572,225

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/866,650
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bennett, J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296, 93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-08-866-650-3

```

Query Match 5.2%; Score 139; DB 2; Length 1013;
Best Local Similarity 26.5%; Pred. No. 3.9e-05;
Matches 39; Conservative 26; Mismatches 48; Indels 34; Gaps 6;

QY 41 SGTMTSKNYPGTYPTVCEKIITVPGKRLRLRLGDLNIES-KTCASDYL-LFSSATDQ 98
DB 782 SGLITSPMWPDKYPSRKCTWVISAIPGHRITLAFNEFEVQHOCAVDHLEIFDGETEK 841
QY 99 ---YGPYCGSNAVPKRELNSNEVTLFKSGSHISGRGFLTYAS----- 140
DB 842 SPILGRLOGS-KIPDPLMATGNEMFIRISDASVQRKGFOATHSTECGRLKAESKPRDL 900
QY 141 -----SDHPDLITC-----LERGS 154
DB 901 YSHAQFGDNNTPGQDCEMLVSEKGS 927

RESULT 6
US-09-021-287-3
Sequence 3, Application US/09021287
Patent No. 5981717
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian tollold-like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/021,287
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/866,650
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296.93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-9166
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-021-287-3

Query Match 5.2%; Score 139; DB 2; Length 1013;
Best Local Similarity 26.5%; Pred. No. 3.9e-05;
Matches 39; Conservative 26; Mismatches 48; Indels 34; Gaps 6;

QY 41 SGTMTSKNYPGTYPTVCEKIITVPGKRLRLRLGDLNIES-KTCASDYL-LFSSATDQ 98
DB 782 SGLITSPMWPDKYPSRKCTWVISAIPGHRITLAFNEFEVQHOCAVDHLEIFDGETEK 841
QY 99 ---YGPYCGSNAVPKRELNSNEVTLFKSGSHISGRGFLTYAS----- 140
DB 842 SPILGRLOGS-KIPDPLMATGNEMFIRISDASVQRKGFOATHSTECGRLKAESKPRDL 900

QY 141 -----SDHPDLITC-----LERGS 154
DB 901 YSHAQFGDNNTPGQDCEMLVSEKGS 927

RESULT 7
US-09-240-473-3
Sequence 3, Application US/09240473
Patent No. 6297011
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian tollold-like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/240,473
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296.93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-9166
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-240-473-3

Query Match 5.2%; Score 139; DB 4; Length 1013;
Best Local Similarity 26.5%; Pred. No. 3.9e-05;
Matches 39; Conservative 26; Mismatches 48; Indels 34; Gaps 6;

QY 41 SGTMTSKNYPGTYPTVCEKIITVPGKRLRLRLGDLNIES-KTCASDYL-LFSSATDQ 98
DB 782 SGLITSPMWPDKYPSRKCTWVISAIPGHRITLAFNEFEVQHOCAVDHLEIFDGETEK 841
QY 99 ---YGPYCGSNAVPKRELNSNEVTLFKSGSHISGRGFLTYAS----- 140
DB 842 SPILGRLOGS-KIPDPLMATGNEMFIRISDASVQRKGFOATHSTECGRLKAESKPRDL 900
QY 141 -----SDHPDLITC-----LERGS 154
DB 901 YSHAQFGDNNTPGQDCEMLVSEKGS 927

RESULT 8
US-08-839-008-9
Sequence 9, Application US/08839008
Patent No. 5916758
GENERAL INFORMATION:
APPLICANT: Hurle, Mark R
APPLICANT: McDonnell, Peter C
APPLICANT: McNulty, Dean E
APPLICANT: Rosen, Craig A
APPLICANT: Siemens, Ivo R

APPLICANT: Young, Peter R
APPLICANT: Yue, Tian-Li
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-9

Query Match 5.1%; Score 137; DB 2; Length 449;

Best Local Similarity 24.8%; Pred. No. 1.6e-05; Matches 62; Conservative 32; Mismatches 98; Indels 58; Gaps 11;

QY 5 AGGSPVALLEFAVCAPRLQAEELGDS-----CGHIVTSODSGTMTSKNYPGTYPTV 58
DB 4 AATASLGLPLTACALLPFAOGQTPNTRPVFLCGDVGK-ESGYVASEGFPNLYPPNKE 62
QY 59 CEKTIIVPKGRRLIL--RLGDLNIESKTCASDYL-LFS---SATDQGYPCGSNAVPKEL 112
DB 63 CIWITIVPEGOVTSLSFRVDELEH--PACRYDALEVPAGSGTSGQRLGRFCGFR-PAPL 120
QY 113 RLNSNEVTVLFKSGSHSGFLITVA-----SSDH 143
DB 121 VAPGNQVTLMTTDEGTGGRFLMTYSGRATSGSEHQFCGRLKAGCTLTTPNMPESDY 180
QY 144 PDLITCLERSHYE-----EKYSKFCPACGR-DIAGDSGNTKGYRDTSLCK 192
DB 181 PPGISCC---SMHTIAPDQVIALTFEKFDELPDTCRYDSVSYFNGAVSDSRLGRFCG 237
QY 193 AAIHAGITTD 202
DB 238 DAVPGSISSE 247

RESULT 9
US-08-839-008-2
; Sequence 2, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:
; APPLICANT: Hurle, Mark R
; APPLICANT: McDonnell, Peter C
; APPLICANT: McNulty, Dean E
; APPLICANT: Rosen, Craig A

APPLICANT: Siemens, Ivo R
APPLICANT: Young, Peter R
TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,008
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/563,697
FILING DATE: 28-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50384
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 449 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-008-2

Query Match 5.1%; Score 135; DB 2; Length 449;

Best Local Similarity 24.8%; Pred. No. 2.6e-05; Matches 62; Conservative 31; Mismatches 99; Indels 58; Gaps 11;

QY 5 AGGSPVALLEFAVCAPRLQAEELGDS-----CGHIVTSODSGTMTSKNYPGTYPTV 58
DB 4 AATASLGLPLTACALLPFAOGQTPNTRPVFLCGDVGK-ESGYVASEGFPNLYPPNKE 62
QY 59 CEKTIIVPKGRRLIL--RLGDLNIESKTCASDYL-LFS---SATDQGYPCGSNAVPKEL 112
DB 63 CIWITIVPEGOVTSLSFRVDELEH--PACRYDALEVPAGSGTSGQRLGRFCGFR-PAPL 120
QY 113 RLNSNEVTVLFKSGSHSGFLITVA-----SSDH 143
DB 121 VAPGNQVTLMTTDEGTGGRFLMTYSGRATSGSEHQFCGRLKAGCTLTTPNMPESDY 180
QY 144 PDLITCLERSHYE-----EKYSKFCPACGR-DIAGDSGNTKGYRDTSLCK 192
DB 181 PPGISCC---SMHTIAPDQVIALTFEKFDELPDTCRYDSVSYFNGAVSDSRLGRFCG 237
QY 193 AAIHAGITTD 202
DB 238 DAVPGSISSE 247

RESULT 10
US-08-991-408-4
; Sequence 4, Application US/08991408
; Patent No. 6008017
; GENERAL INFORMATION:
; APPLICANT: ARLETH, ANTHONY J.
; APPLICANT: WILLETTTE, ROBERT N.
; APPLICANT: ELSHOUBAGY, NABIL A.

APPLICANT: LI, XIAOTONG
TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLLOID-LIKE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,408
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/034,471
FILING DATE: 02-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50038
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 591 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-408-4

Query Match 5.1%; Score 135; DB 3; Length 591;
Best Local Similarity 25.2%; Pred. No. 4.1e-05;
Matches 53; Conservative 34; Mismatches 73; Indels 50; Gaps 12;

QY 18 CAPRLQAEELD-----GCGHIVTSODSGTMTSKNPGTYPNVTCERKITVPGKGR 70
DB 179 CEP-----GYELGPPDRRSEACGGLITKL-NGTITTPGPKKEYPPNKNCWQVVAPTQYR 233
QY 71 LILRLGDLNIE-SKTCASDYLL-----FSSATDQYGYCGSMAYPKRELRLNSNEVTVLFKS 125
DB 234 ISVKKEFFELNEGVCKDYVEIMSGLSSESKLHGKFCGA-EVPEVITSQFNMMRIEFS 292
QY 126 GSHISGRGFLTYASSDHPDLITCLERGSHYEEK--YSK---FCPAGCRDIAGDISGNT 180
DB 293 DNTVSKKGF-----KAHFFSDKDECKNDGGQHECVNMGSYMCC 334
QY 181 KDGY--RDTSLCKAA-----IH--AGITT 201
DB 335 RNFGLVLDNKKHDKKEACEQKIHSPSGILT 364

RESULT 11
US-09-432-473-4
Sequence 4, Application US/09432473
Patent No. 6365715
GENERAL INFORMATION:
APPLICANT: ARLETH, ANTHONY J.
APPLICANT: WILBITE, ROBERT N.
APPLICANT: ELSHOUBAGY, NABIL A.
APPLICANT: LI, XIAOTONG
TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOLLLOID-LIKE PROTEIN
FILE REFERENCE: ATG-50038-DI
CURRENT APPLICATION NUMBER: US/09/432,473

CURRENT FILING DATE: 1999-11-01
EARLIER APPLICATION NUMBER: 08/991,408
EARLIER FILING DATE: 1997-12-16
EARLIER APPLICATION NUMBER: 60/034,471
EARLIER FILING DATE: 1997-01-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 591
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-432-473-4

Query Match 5.1%; Score 135; DB 4; Length 591;
Best Local Similarity 25.2%; Pred. No. 4.1e-05;
Matches 53; Conservative 34; Mismatches 73; Indels 50; Gaps 12;

QY 18 CAPRLQAEELD-----GCGHIVTSODSGTMTSKNPGTYPNVTCERKITVPGKGR 70
DB 179 CEP-----GYELGPPDRRSEACGGLITKL-NGTITTPGPKKEYPPNKNCWQVVAPTQYR 233
QY 71 LILRLGDLNIE-SKTCASDYLL-----FSSATDQYGYCGSMAYPKRELRLNSNEVTVLFKS 125
DB 234 ISVKKEFFELNEGVCKDYVEIMSGLSSESKLHGKFCGA-EVPEVITSQFNMMRIEFS 292
QY 126 GSHISGRGFLTYASSDHPDLITCLERGSHYEEK--YSK---FCPAGCRDIAGDISGNT 180
DB 293 DNTVSKKGF-----KAHFFSDKDECKNDGGQHECVNMGSYMCC 334
QY 181 KDGY--RDTSLCKAA-----IH--AGITT 201
DB 335 RNFGLVLDNKKHDKKEACEQKIHSPSGILT 364

RESULT 12
US-08-866-650-5
Sequence 5, Application US/0886650
Patent No. 5939321
GENERAL INFORMATION:
APPLICANT: Greenspan, Daniel S
APPLICANT: Takahara, Kazuhiko
APPLICANT: Hoffman, Guy G
TITLE OF INVENTION: Mammalian Tollloid-Like Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pluckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/866,650
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Beison, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 960296,93839
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-866-650-5

Query Match	5.1%;	Score 135;	DB 2;	Length 1013;
Best Local Similarity	23.1%;	Pred. NO. 0.0001;		
Matches	42;	Conservative	31;	Mismatches 67;
				Indels 42;
				Gaps 8;

```

QY      32  CGHIVTSODSGTMTSKNVPGPYPPYATYCEKIIYVPRKRLILRAGDNTI--ESTQCSADYL  90
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      349  CGEIL--QESNQNLSPPGFPMQIPSYTHCIRKVSATYTPBEKIVLANFTTMDLYKSSLCMTDI  407

QY      91  LESSATDOY-----GPYCGSMAVPKRELRLNSNEVTLFKSGSHISGRGFLITYASSDH  143
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      408  ---EVRDGYWKRKSPDLGRFCGD--KLPEVLNSTDSBMMIEPRSSSNWVGKGFAYVEA---  460

QY      144  PDLITLCBERGSHYEKEYSKRCFPCAGCDIAGDI--SGTTCGQIRDT-----SLICKAAITHA  197
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      461  -----ICGGEIRKKNQGIQSPPYPPDYRPMKCEVWKITVSESEYHV  500

QY      198  GI  199
      | :
Db      501  GL  502

```

RESULT 13

```

US-09-021-287-5
: Sequence 55, Application US/09021287
: Patent No. 5981717
:
: GENERAL INFORMATION:
:
: APPLICANT: Greenspan, Daniel S
:
: APPLICANT: Takahata, Kazuhiko
:
: APPLICANT: Hoffman, Guy G
:
: TITLE OF INVENTION: Mammalian Tolloid-Like Protein
:
: NUMBER OF SEQUENCES: 13
:
: CORRESPONDENCE ADDRESSES:
:
: ADDRESSEE: Quarles & Brady
:
: STREET: 1 South Pinckney Street
:
: CITY: Madison
:
: STATE: WI
:
: COUNTRY: US
:
: ZIP: 53703
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION NUMBER:
:
: APPLICATION NUMBER: US/09/021,287
:
: FILING DATE:
:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: 08/866,650
:
: FILING DATE:
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Berson, Bennett J
:
: REGISTRATION NUMBER: 37094
:
: REFERENCE/DOCKET NUMBER: 960296.93839
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 608-251-5000
:
: TELEFAX: 608-251-9166
:
: INFORMATION FOR SEQ ID NO: 5:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1013 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-09-021-287-5

```

```

Query Match      5.1%  Score 135;  DB 2;  Length 1013;
Best Local Similarity 23.1%  Pred. No. 0.0001;
Matches 42;  Conservative 31;  Mismatches 67;  Indels 42;  Gaps 8;

OY      32 CGHVTSDSGTMSKNTPGTYPNNTVEKITTVEKGRRLIIRLGDLMN-ESKTCASDYL 90
      || : : : : : : || : || : : : : : : : : : : : : : : : : : || :

```

```

Db      349 CGETL-QESNGNLSSPGFPNMGPSYTHCIWRVSVTPGEKIVLNETTMDLYKSSLCMYDI 407
QY      91 LFSSATQY-----GPGGSAVVPKELRLNSNEVTVLEKSGSHISGRFLLTVASSDH 143
           |         |         |         |         |         |         |
Db      408 ---EVRDGYWRKSPLLGRFCGD-KLPEVLITSDSRMWIEFRSSSNWYGGKFAAVYE--- 460
QY      144 PDLITCLERGSHYEERKSKFPCAGCDIAGDI-SGNTKGYRDT-----SLCKAIIHA 197
           |         |         |         |         |         |         |
Db      461 -----ICGGEIRKNEGQIQSPNYPDPDYRMKCYWKITVSESEYHV 500
QY      198 GI 199
           |         |
Db      501 GL 502

RESULT 14
#F:08-001-408-3

```

RESULT 14

```

US-08-991-408-2
Sequence 2, Application us/08991408
Patent No. 6008017
GENERAL INFORMATION:
APPLICANT: AILETH, ANTHONY J.
APPLICANT: WILLETTTE, ROBERT N.
APPLICANT: ELSHOURBAGY, NABIL A.
APPLICANT: LI, XIAOMING
TITLE OF INVENTION: HUMAN CARDIAC/BRAIN TOL
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: RAYNER & PRESTIA
STREET: P.O. BOX 960
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0.
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,408
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/034,471
FILING DATE: 02-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50038
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1013 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-408-2

```

```

Query Match      5.1%: Score 135; DB 3; Length 1013;
Best Local Similarity 23.1%: Pred. NO. 0.0001:
Matches 42; Conservative 31; Mismatches 67; Indels 42; Gaps 8;

QY      32 CGHIVTSQDSGTMSTKNPGYTPNNTVCCEKITIVYKGRRLILRLDLNI-ESEKTCASDYL 90
      || : : : : : || : : : : : || : : : : : || : : : : : || :
Db      349 CGEFL-QESNNLSPPGNGCPYSTTHICIMRWSVIPGKRIYVNFMTMLDKYSLCLMYDI 407
      || : : : : : || : : : : : || : : : : : || : : : : : || :

QY      91 LFSSATDY-----GPGCGMAVPEKELRLNLSNEVYLFKSGSHISGRGFLITYASSDH 143
      || : : : : : || : : : : : || : : : : : || : : : : : || :

```

Search completed: May 15, 2003, 13:15:23
Job time : 16.5632 secs

Db 408 ---EVRDGYWRKSPILGRPCGD-KLPEVLTSTDSNMWIEFRSSNMWVGKGFAYVEA--- 460
QY 144 PDLITCLERGSHYFEKYSKFCPACGRDIAGDI-SGNTKDGYPD-----SLICKAAIHA 197
Db 461 -----ICGGEIRKNEGQIOSPNYPDDYRPMKECYWKITVSESYHV 500
QY 198 GI 199
Db 501 GI 502

RESULT 15

US-09-240-473-5
; Sequence 5, Application US/09240473
; Patent No. 6297011
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Takahara, Kazuhiko
; APPLICANT: Hoffman, Guy G
; TITLE OF INVENTION: Mammalian Tolloid-Like Protein
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/240.473
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Berson, Bennett J
; REGISTRATION NUMBER: 37094
; REFERENCE/DOCKET NUMBER: 960296.93839
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1013 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-240-473-5

Query Match 5.1%; Score 135; DB 4; Length 1013;
Best Local Similarity 23.1%; Pred. No. 0.0001;
Matches 42; Conservative 31; Mismatches 67; Indels 42; Gaps 8;

QY 32 CGHIVTSDSGTMTSKNPGTYPNVCEKILTVPKRLILRLGLDINT-ESKTCASDYL 90
Db 349 CGEYL-QESNGNLSPGPNPGYTHCIWRVSYPERKIVLFTTMDLYKSSLCWYDI 407
QY 91 LESSATDOY-----GYCCSMAYPKELRLNSNEYVLFKSGSHISGKGFLLTYAASDH 143
Db 408 ---EVRDGYWRKSPILGRPCGD-KLPEVLTSTDSNMWIEFRSSNMWVGKGFAYVEA--- 460
QY 144 PDLITCLERGSHYFEKYSKFCPACGRDIAGDI-SGNTKDGYPD-----SLICKAAIHA 197
Db 461 -----ICGGEIRKNEGQIOSPNYPDDYRPMKECYWKITVSESYHV 500
QY 198 GI 199
Db 501 GI 502

QY 241 FTTPGMNTTVAIPSVIFIALLLTGMGIFALICRRKKKGNPNVSADAKTCGCMQIKYF 300
|||||
Db 241 FTTPGMNTTVAIPSVIFIALLLTGMGIFALICRRKKKGNPNVSADAKTCGCMQIKYF 300
QY 301 ARHOSTEFTIYDNKEKMTOKLDLITSDMADYQOPLMGTGTVAARKSTFFRPMOTDEEV 360
|||||
Db 301 ARHOSTEFTIYDNKEKMTOKLDLITSDMADYQOPLMGTGTVAARKSTFFRPMOTDEEV 360
QY 361 RVNTEASGHYCCPHRPGHEHEALPLTHSEPEYATPIVERHLLRAHTFSTOGSYRVPGPRP 420
|||||
Db 361 RVNTEASGHYCCPHRPGHEHEALPLTHSEPEYATPIVERHLLRAHTFSTOGSYRVPGPRP 420
QY 421 THKSHSSGPPATGATQVESYORPASPKPVGGYDKPAASSFLDSRDPASQSGMTSGG 480
|||||
Db 421 THKSHSSGPPATGATQVESYORPASPKPVGGYDKPAASSFLDSRDPASQSGMTSGG 480
QY 481 DDGYSAPRNGIAPLNOTAMTALL 503
|||||
Db 481 DDGYSAPRNGIAPLNOTAMTALL 503

RESULT 2

US-09-823-038A-51
; Sequence 51, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevlin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumbie, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1037C3
; CURRENT APPLICATION NUMBER: US/09/823.038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 503
; TYPE: PRP
; ORGANISM: Mouse
US-09-823-038A-51

Query Match 99.6%; Score 2659; DB 10; Length 503;
Best Local Similarity 99.6%; Pred. No. 4.1e-209;
Matches 501; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGTAGGSPVIALLFVAVCAPRLQAEELGDCGHIIVTSQDSGTMTSKNYPGTYPNYTVC 60
|||||
Db 1 MGTAGGSPVIALLFVAVCAPRLQAEELGDCGHIIVTSQDSGTMTSKNYPGTYPNYTVC 60
QY 61 KIITVPKGRLLRLRGDNIESTKTCASDYLLFSSATDQGYPCGSAWPKKELRLNSNEVT 120
|||||
Db 61 KIITVPKGRLLRLRGDNIESTKTCASDYLLFSSATDQGYPCGSAWPKKELRLNSNEVT 120
QY 121 VLFKSGSHISGRGFLITVASSDHPDLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 180
|||||
Db 121 VLFKSGSHISGRGFLITVASSDHPDLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 180
QY 121 VLFKSGSHISGRGFLITVASSDHPDLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 180
|||||
Db 121 VLFKSGSHISGRGFLITVASSDHPDLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 180
QY 181 KDGYRDTSLCKAAIHAGIITDELGGINLLQSKGISHYEGLANGVLSRHGSLSEKREFL 240
|||||
Db 181 KDGYRDTSLCKAAIHAGIITDELGGINLLQSKGISHYEGLANGVLSRHGSLSEKREFL 240
QY 241 FTTPGMNTTVAIPSVIFIALLLTGMGIFALICRRKKKGNPNVSADAKTCGCMQIKYF 300
|||||
Db 241 FTTPGMNTTVAIPSVIFIALLLTGMGIFALICRRKKKGNPNVSADAKTCGCMQIKYF 300
QY 301 ARHOSTEFTIYDNKEKMTOKLDLITSDMADYQOPLMGTGTVAARKSTFFRPMOTDEEV 360
|||||
Db 301 ARHOSTEFTIYDNKEKMTOKLDLITSDMADYQOPLMGTGTVAARKSTFFRPMOTDEEV 360

QY 361 RVNTEASGHYCCPHRPGHEHEALPLTHSEPEYATPIVERHLLRAHTFSTOGSYRVPGPRP 420
|||||
Db 361 RVNTEASGHYCCPHRPGHEHEALPLTHSEPEYATPIVERHLLRAHTFSTOGSYRVPGPRP 420
QY 421 THKSHSSGPPATGATQVESYORPASPKPVGGYDKPAASSFLDSRDPASQSGMTSGG 480
|||||
Db 421 THKSHSSGPPATGATQVESYORPASPKPVGGYDKPAASSFLDSRDPASQSGMTSGG 480
QY 481 DDGYSAPRNGIAPLNOTAMTALL 503
|||||
Db 481 DDGYSAPRNGIAPLNOTAMTALL 503

RESULT 3

US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003.132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRP
; ORGANISM: Mus musculus
US-10-003-132-6

Query Match 89.7%; Score 2395.5; DB 9; Length 458;
Best Local Similarity 91.1%; Pred. No. 1.2e-187;
Matches 458; Conservative 0; Mismatches 0; Indels 45; Gaps 1;

QY 1 MGTAGGSPVIALLFVAVCAPRLQAEELGDCGHIIVTSQDSGTMTSKNYPGTYPNYTVC 60
|||||
Db 1 MGTAGGSPVIALLFVAVCAPRLQAEELGDCGHIIVTSQDSGTMTSKNYPGTYPNYTVC 60
QY 61 KIITVPKGRLLRLRGDNIESTKTCASDYLLFSSATDQGYPCGSAWPKKELRLNSNEVT 120
|||||
Db 61 KIITVPKGRLLRLRGDNIESTKTCASDYLLFSSATDQGYPCGSAWPKKELRLNSNEVT 120
QY 121 VLFKSGSHISGRGFLITVASSDHPDLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 180
|||||
Db 121 VLFKSGSHISGRGFLITVASSDHPDLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 180
QY 100 -----DLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 135
|||||
Db 100 -----DLITCLERGSHYFEKYSKPCPACRDIAGDISGNT 135
QY 181 KDGYRDTSLCKAAIHAGIITDELGGINLLQSKGISHYEGLANGVLSRHGSLSEKREFL 240
|||||
Db 181 KDGYRDTSLCKAAIHAGIITDELGGINLLQSKGISHYEGLANGVLSRHGSLSEKREFL 240
QY 241 FTTPGMNTTVAIPSVIFIALLLTGMGIFALICRRKKKGNPNVSADAKTCGCMQIKYF 300
|||||
Db 241 FTTPGMNTTVAIPSVIFIALLLTGMGIFALICRRKKKGNPNVSADAKTCGCMQIKYF 300
QY 301 ARHOSTEFTIYDNKEKMTOKLDLITSDMADYQOPLMGTGTVAARKSTFFRPMOTDEEV 360
|||||
Db 301 ARHOSTEFTIYDNKEKMTOKLDLITSDMADYQOPLMGTGTVAARKSTFFRPMOTDEEV 360
QY 361 RVNTEASGHYCCPHRPGHEHEALPLTHSEPEYATPIVERHLLRAHTFSTOGSYRVPGPRP 420
|||||
Db 361 RVNTEASGHYCCPHRPGHEHEALPLTHSEPEYATPIVERHLLRAHTFSTOGSYRVPGPRP 420
QY 421 THKSHSSGPPATGATQVESYORPASPKPVGGYDKPAASSFLDSRDPASQSGMTSGG 480
|||||
Db 421 THKSHSSGPPATGATQVESYORPASPKPVGGYDKPAASSFLDSRDPASQSGMTSGG 480
QY 481 DDGYSAPRNGIAPLNOTAMTALL 503
|||||

```

DB          436 DDGYSAPRNGLAPLNQNTAMTALL 458

RESULT 4
US-10-003-132-2
; Sequence 2, Application US/10003332
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROKILLIN HOMOLOG ZCDB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-003-132-2

```

Query Match	72.1%;	Score 1924.5;	DB 9;	Length 715;
Best Local Similarity	55.8%;	Pred. No. 7.3e-149;		
Matches 392;	Conservative 34;	Mismatches 73;	Indels 203;	Gaps 4

OY	5	AGGPEVLLPAAVCAPLELAEEGLDGGCHIVTSQDSGTMKSKYPCPYIPVYTCETIT	64
Db	14	AAGRLALALLAASAPRLQAELEGLDGGCHLYTVQDSGTMKSKYPCPYIRNHVYCEITIT	73
OY	65	VPKGRRLILRLGDLNIESKTCASDYLLFSSATDQYGPYCGSWAPKRELUNSNVTVLFK	124
Db	74	VPKGRRLILRLGDLNIESQTCASDYLLFTSSDDQYGPYCGSMVPEKELLANTSEVTVFE	133
OY	125	SGSHSGGFLITTAASDPDLITCLERGHSHYFEKSKPCNCCRIADISGNTDGY	184
Db	134	SGSHSGGFLITTAASDPDLITCLERASHYLTETSKPCNCCRIADISGNTDGY	193
OY	185	RDTSILCKAAIHAGIITDELGHINLQSKGISHYEGILANGVLSRRGSLSEKRFLE	241
Db	194	RDTSILCKAAIHAGIITDELGHINLQSKGISHYEGILANGVLSRDSLSDKRFLETSN	253
OY	242	-----241-----	
Db	254	GCSRSLSFEPDQIRASSSGWSQSVNESGDQYHMSPGQARLQDQGPSMASGDSNNHKREW	313
OY	242	-----241-----	
Db	314	LEIDLGKKKITTGIRTGTSQTSQSNFVYKSVNMFKNKNSKMTYKCIYNNEKVFQDGS	373
OY	242	-----241-----	
Db	374	NFRDPVQNNFIPPIVARYRVVPQTHQRIALKYELIGCOITQGNDSLVRKTSQTSYS	433
OY	242	-----TPGKNITVVALPSVIFALLITMGIFALICRKKKKGNPVYA	285
Db	434	TKKEDETTRIPPEEIESTGINTITVVALPLVLVAFAGGIFPAARFKRKKGSPGSA	493
OY	286	DAQTKGCKQIKTPEPARHOSIEFTISYDNKEKMTOKDLITSDMADYQOPLMIGTGVAR	345
Db	494	EAQTKGCKQIKTPEPARHOSIEFTISYDNKEKMTOKDLITSDMADYQOPLMIGTGVYR	553
OY	346	KGSTFRPMDDTTEEVRYNTEASGHYDCPHRPGHEHYALPLTHSEPEYATPIVERHLLRAH	405
Db	554	KGSTFRPMDDTTEAEAGVSTDAAGHYDCPORAGRHEHYALPLAPEPEYATPIVERHYLRAH	613
OY	406	TFSTOSGGRVYGPPTKHSHSSGSGPPRANG-ATQVESYORPASPKRVGGGYOKRPASSP	464
Db	614	TFSMOSGGRVYGPQGHKSHSSGSGFSPVAGVCAQDQYORPHSAQPADRKQIDPRAVSA	673

```

QY      465 L-DSRDPISQSQMT-SGDDGGSAPRNGLPLNQMTALL 503
      : | | | | | | | | | | | | | | | | | | | |
Db      674 LATESGHPSQKPTTHPGNSDSYSAPRDCLPLNQMTALL 715

```

```

1 RESULT 5
2 US-09-759-130B-73
3 Publication 73, Application US/09759130B
4 Publication No. US20030022279A1
5 GENERAL INFORMATION:
6 APPLICANT: Millennium Pharmaceuticals, Inc.
7 APPLICANT: McCarthy, Sean A
8 APPLICANT: Fraser, Christopher C
9 APPLICANT: Sharp, John D
10 APPLICANT: Barnes, Thomas S
11 APPLICANT: Kirsch, Susan J
12 APPLICANT: Mackay, Charles R
13 APPLICANT: Myers, Paul S
14 APPLICANT: Leiby, Kevin R
15 APPLICANT: Wrighton, Nicolas
16 APPLICANT: Goodearl, Andrew
17 APPLICANT: Holtzman, Douglas A
18 TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
19 TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
20 TITLE OF INVENTION: USES.
21 FILE REFERENCE: MP100-535ONMIM
22 CURRENT APPLICATION NUMBER: US/09/759,130B
23 CURRENT FILING DATE: 2002-09-16
24 PRIOR APPLICATION NUMBER: US 09/479,249
25 PRIOR FILING DATE: 2000-01-07
26 PRIOR APPLICATION NUMBER: US 09/559,497
27 PRIOR FILING DATE: 2000-04-27
28 PRIOR APPLICATION NUMBER: US 09/578,063
29 PRIOR FILING DATE: 2000-05-24
30 PRIOR APPLICATION NUMBER: US 09/333,159
31 PRIOR FILING DATE: 1999-06-14
32 PRIOR APPLICATION NUMBER: US 09/596,194
33 PRIOR FILING DATE: 2000-07-14
34 PRIOR APPLICATION NUMBER: US 09/342,364
35 PRIOR FILING DATE: 1999-06-29
36 PRIOR APPLICATION NUMBER: US 09/608,452
37 PRIOR FILING DATE: 2000-06-30
38 PRIOR APPLICATION NUMBER: US 09/393,996
39 PRIOR FILING DATE: 1999-09-10
40 PRIOR APPLICATION NUMBER: US 09/602,871
41 PRIOR FILING DATE: 2000-06-23
42 PRIOR APPLICATION NUMBER: US 09/420,707
43 PRIOR FILING DATE: 1999-10-19
44 NUMBER OF SEQ ID NOS: 460
45 SOFTWARE: FastSeq for Windows Version 4.0
46 SEQ ID NO 73
47 LENGTH: 715
48 TYPE: prt
49 ORGANISM: Homo sapiens
50 US-09-759-130B-73

```

[illegible]

```
|||||
Db 194 RPTSLKCAAHAGIITADELGGQISVLQKRGISREGLIANGVLRDGLSKRFLTSN 253
Qy 242 ----- 241
Db 254 GCSRLSEFPDQIIRASSWSQVNSGDOVHMSPGQARLQDQSPMASGSSNNHKPREM 313
Qy 242 ----- 241
Db 314 LEIDGERKKITGIRTYGSTQSNFNFVYKSFVNFKNNSKWKTKGIYNNEKVFQGNS 373
Qy 242 ----- 241
Db 374 NFRDPOVNNFPIYARVYRVVPTWQHRIALKVELIGCQITQGNDLSLWKRKTSSTYS 433
Qy 242 -----TPGNNITVAIPSVIFIALLTGMCIFAICRRKKRKNPYSA 285
Db 434 TKKEDETIRPISEETSGINITVAIPVLVYLVFAGMGIFAFRKKKKSGPSYGA 493
Qy 286 DAQKTCGKQIKYPPARHOSSTFTSYDNEKEMTKLDLITSDMADYQOPLMIGTVAR 345
Db 494 EAKQKDCWKQIKYPPARHOSAEFTSYDNEKEMTKLDLITSDMADYQOPLMIGTVAR 553
Qy 346 KGSTFRPMDTDEEYRVNTEASGHDCPHRGRHEVALPLTSEPEYATPIYERHLLRAH 405
Db 554 KGSTFRPMDTDEEYRVNTEASGHDCPHRGRHEVALPLTSEPEYATPIYERHLLRAH 613
Qy 406 TFSSTGGRVPPRPTTHKSHSSGSGFPATG-ATQVESYQASPRKPYGGYDKPAASFS 464
Db 614 TFSAGSGRVPPGPHKSHSSGSGFPATGAGDQYGRHSAQPADRGYDRKAVSA 673
Qy 465 L-DSRDPASQSQMT-SGDDQYSAPRNGLAPLNOTAMTALL 503
Db 674 LATESGHPDSQKPPTHPGHSDYSAPRDCLTPLNOTAMTALL 715

RESULT 6
US-09-759-130B-75
; Sequence 75, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: McCarty, Sean A
; APPLICANT: Fraser, Christopher C
; APPLICANT: Sharp, John D
; APPLICANT: Barnes, Thomas S
; APPLICANT: Kirt, Susan J
; APPLICANT: Mackay, Charles R
; APPLICANT: Myers, Paul S
; APPLICANT: Leiby, Kevin R
; APPLICANT: Wrighton, Nicolas
; APPLICANT: Goodheart, Andrew
; APPLICANT: Holtzman, Douglas A
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: MP100-5350M1M
; CURRENT APPLICATION NUMBER: US/09/759,130B
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 09/479,249
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: US 09/559,497
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/578,063
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: US 09/333,159
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 09/596,194
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/342,364
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 09/608,452
; PRIOR FILING DATE: 2000-06-30
```

```
; PRIOR APPLICATION NUMBER: US 09/393,996
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: US 09/602,871
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/420,707
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 681
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-130B-75

Query Match 69.6%; Score 1859.5; DB 9; Length 681;
Best Local Similarity 55.4%; Pred. No. 1,4e-143;
Matches 377; Conservative 33; Mismatches 68; Indels 203; Gaps 4;

Qy 26 BELGDCGCHVTSDSGTMSKNYPTGTYPERKILTVPKKLLRLGDLNIESKTC 85
Db 1 BELGDCGCHVTYODSGTMSKNYPTGTYPERKILTVPKKLLRLGDLNIESQTC 60
Qy 86 ASDYLLFSATDQYPPGSGMAVPELRNLSNEVTVLPFGSGSHISGRGLTYASSDHPD 145
Db 61 ASDYLLFTSSDQYPPGSGMTVPKELNLTSEVTVRESGSHISGRGLTYASSDHPD 120
Qy 146 LITCLERSHYEERKSKPCACRDIAGDISGNKRGYRDTSLCKAIAHAGITDELG 205
Db 121 LITCLERSHYEERKSKPCACRDIAGDISGNKRGYRDTSLCKAIAHAGITDELG 180
Qy 206 GHINLQSGKISHYEGGLANGVLSRHSSEKRFLE----- 241
Db 181 GQISVLQKRGISRYEGIIANGVLSRHSSEKRFLE----- 240
Qy 242 ----- 241
Db 241 SYNESGDOVHMSPGQARLQDQSPMASGSSNNHKPREMLEIDLEKKITGIRTTGSIQ 300
Qy 242 ----- 241
Db 301 SNFNFYKSFVNFKNNSKWKTKGIYNNEKVFQGNSNFRDPVONNPIPIVARYRV 360
Qy 242 -----TTGCM 246
Db 361 VPQTHQRIALKVELIGCQITQGNDLSLWKRKTSSTSTKREDETTTRPISEETSGI 420
Qy 247 NITVAIPSVIFIALLTGMCIFAICRRKKRKNPYSAQKTCGKQIKYPPARHOS 306
Db 421 NITVAIPVLVYLVFAGMGIFAFRKKKKSGPSYGAENQKTCGKQIKYPPARHOSA 480
Qy 307 EFTISYDNEKEMTKLDLITSDMADYQOPLMIGTVARKSGTFRPMDTDEEYRVNTEA 366
Db 481 EFTISYDNEKEMTKLDLITSDMADYQOPLMIGTVARKSGTFRPMDTDEEYRVNTEA 540
Qy 367 SGHYCCPHRPGHHEVALPLTSEPEYATPIYERHLLRAHTEFTOGGYRVGGRPHKHSH 426
Db 541 GGHYCCPHRPGHHEVALPLTSEPEYATPIYERHLLRAHTEFTOGGYRVGGRPHKHSH 600
Qy 427 SSGGFPATG-ATQVESYQASPRKPYGGYDKPAASFL--DSRDPASQSQMT-SGDD 482
Db 601 SSGGFPATG-ATQVESYQASPRKPYGGYDKPAASFL--DSRDPASQSQMT-SGDD 482
Qy 483 GYSAPRNGIAPLNOTAMTALL 503
Db 661 GYSAPRNGIAPLNOTAMTALL 681

RESULT 7
US-09-759-130B-76
; Sequence 76, Application US/09759130B
; Publication No. US2003002279A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
```


RESULT 8
US-09-759-130B-78
: Sequence 78, Application US/09759130B
: Publication No. US20030022279A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc
: APPLICANT: McCarthy, Sean A

```

RESULT 9 -
US-10-060-830-3
Sequence 3, Application US/10060830
Publication No. US2003003254A1
GENERAL INFORMATION:
APPLICANT: Gu, Yizhong
INVENTOR: Neuyen, Cung-Tuong
TITLE OF INVENTION: HUMAN ICCL DOMAIN CONTAINING PROTEIN

```

FILE REFERENCE: PB0169
 CURRENT APPLICATION NUMBER: US/10/060, 830
 CURRENT FILING DATE: 2002-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 09/864, 761
 PRIOR FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/325, 062
 PRIOR FILING DATE: 2001-09-25
 NUMBER OF SEQ ID NOS: 1123
 SOFTWARE: Neomica Sequence Listing Engine
 SEQ ID NO 3
 LENGTH: 729
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-060-830-3

Query Match 21.5%; Score 575.5; DB 9; Length 729;
 Best Local Similarity 24.5%; Pred. No. 1.4e-38;
 Matches 183; Conservative 84; Mismatches 181; Indels 299; Gaps 20;

QY 8 PSVALLEAVCAPRLQAEELGDCGGHVTSDSGTMTSKNPGTYPNVTCERKITYPK 67
 DB 2 PLFLLLVLLVLLLEDAGAOOGDGGHVTLSPESGTLTINPQTYPNVTCERKITYPK 61
 QY 68 GKRLRLGLDNLIE-SKTCASDYLLFS-ATDQGYPCG-SWAVPKELRLNSENVY 121
 DB 62 GERVAIKGDDIEDSDSCHENYLTNYGIVSKTEIKRYCGLGQNMHSIESKNEITL 121
 QY 122 LFRSGSHISGRGFLTTYSSDHPDLITCLERSHFEERYSKFCPAGCRDIAGDISGNTK 181
 DB 122 LFMGSHVSGRGLASVVIDKODLITCLDPAASNLEPEFSKCPAGCLLPFAELISGTL 181
 QY 182 DGYRPTSLCAAIHAGIITDELGCHINLQSKGISHYEGLLANVLSRHGSLSKRFL 241
 DB 182 HGYSRSPCLMAGVAGVAVSITLGGQISVLSKGIPIYESSLIANNVTSVGHLSLFT 241
 QY 242 TT----- 243
 DB 242 KTSCTCYGLMGESVIADPQITASSVLEWTDHTGOENSMKPKKARKPPMAAFATDE 301
 QY 244 -----PGM----- 246
 DB 302 YQMLQIDLNKREKITGITGTSWEHNYVSAYRIKLYSDQOKWTVYREGVEODKIFQ 361
 QY 247 ----- 246
 DB 362 GNKDYHODVRNNLPPIIARFIRVNPQMOQKIAKMKELLCQGITPKGRPKLTQPPPR 421
 QY 247 -----NIT-TV 251
 DB 422 NSNDLKNTPAKTAKGAPFTQPLQPRSSNEPPAQTEQITASSDINDINTVTPVNTADV 481
 QY 252 AIPSVI-FIALLLTGMGIFAIC-----RRKKKGN-EVVSADAKTCGKWKQKYP 300
 DB 482 ALAVALVAVLVAVLTTLLILVCAWHNRNKKKTEGTYDLP-----WDRAGWKKMGKQFL 537
 QY 301 -----ARHSTFETISYDNKEKMTOKLILITSMDADYQOPLMIGITVARKGST 349
 DB 538 PAKAVDHEETPVRSSE-VNHLSPREVT-----TVLQADSAYAPLIVGTLGHORST 592
 QY 350 FRPMDTDEEVYVMTBAGH-----YDCPHRGRHEVALPLTHSREVAATPIVERHLLRA 404

DB 593 FRP-----EKEGAYADLPYNSBQEVYHAIVAEPLITGPERATPII----- 636
 QY 405 HTFSTQSYRVPG-PRPTHKSHSSGSGPPATGATQVSYORPASPAPKVGGYDKP-AAS 462
 DB 637 ---MMSGHPTTSVGPSTSTFEKATGNQPPPLVGYTNLLSTBDCSSAQAQYDPRKAGK 693
 QY 463 SFLDSRD-----PASQSMTSGCCDG 483
 DB 694 PELPAPDELIVYQVPQSTQEVSGAGRDG 720

RESULT 10

US-10-060-830-1114
 ; Sequence 1114, Application US/10060830
 ; Publication No. US20030032154A1
 ; GENERAL INFORMATION:

APPLICANT: Gu, Yizhong
 APPLICANT: Nguyen, Cung-Tuong
 TITLE OF INVENTION: HUMAN LCCL DOMAN CONTAINING PROTEIN
 FILE REFERENCE: PB0169

CURRENT APPLICATION NUMBER: US/10/060, 830
 CURRENT FILING DATE: 2002-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 09/864, 761
 PRIOR FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/325, 062
 PRIOR FILING DATE: 2001-09-25
 NUMBER OF SEQ ID NOS: 1123
 SOFTWARE: Neomica Sequence Listing Engine
 SEQ ID NO 1114
 LENGTH: 653
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-060-830-1114

Query Match 15.6%; Score 417; DB 9; Length 653;
 Best Local Similarity 22.3%; Pred. No. 1.1e-25;
 Matches 147; Conservative 70; Mismatches 149; Indels 294; Gaps 18;

QY 90 LFFSSATDQGYPCG-SWAVPKELRLNSENVYLFKSGSHISGRGFLTTYASSDHPDLIT 148
 DB 13 LLEDAGAQQKRYCGLGQNMHSIESKNEITLLEFMGSHVSGRGLASVVIDKODLIT 72
 QY 149 CLERSHFEERYSKFCPAGCRDIAGDISGNTKDCYRPTSLCAAIHAGIITDELGHI 208
 DB 73 CLDPAASNLEPEFSKYCPAGCLLPFAELISGTLPHCYRDSPLCMAGVAVSITLGGQI 132
 QY 209 NLQSKGISHYEGLLANVLSRHGSLSEKRELFIT----- 243
 DB 133 SVVLSKGIPIYESSLIANNVTSVGHLSLFTFKTSGCYGLMGESVIADPQITASSVL 192
 QY 244 ----- 243
 DB 193 EWTBDHTGOENSMKPKKARKPPMAAFATDEVQMLQIDLNKREKITGITGTSWEH 252
 QY 244 -----PGM----- 246
 DB 253 NYVSAYRIKLYSDQOKWTVYREGVEODKIFQGNKDYHODVRNNLPPIIARFIRVNP 312
 QY 247 ----- 246
 DB 313 QMOQKIAKMKELLCQGITPKGRPKLTQPPPRNSNDLKNTPAKTAKGAPFTQPLQ 372

```

OY 247 -----NIT-TVAIDSVI--FIALITGICFAIC--- 272
DB 373 PRSSNEPQTEQOTASPDIRNTTVPNTKDVALAVALPVLMTTLILIVCAMW 432
OY 273 --RKRKKGN--PYVSADAKTCCKQKIKYPP-----AARQSTEFISTDNEK 316
DB 433 RNRKKKTEGYDLPY--WDRAGMMKMQFLPARAVDHEETPVRSSE--VNHLSPR 486
OY 317 EMTOKDLITSDMADYQOPLMIGTIVARKGSTFRPMQDTEEVRYNTEASGH-----YD 371
DB 487 EYI---TVLOADSAEYTAQPLVGGIVGTLHQRSTFKP-----EE-----GKAGTADLDPY 534
OY 372 CPHRPGRHEVALPLTHSEPEYATPIVERHLLRAHTEFSTOSGYRVPG--PRPTHRHSHSGG 430
DB 535 SPQGEYHYHAYAEPLITGPEYATPII-----MDMSGHPTTSVGPSTSTFRATGN 584
OY 431 FPPRATATQVESQORASRPVGGYDKP--AASSFLDSRP-----PASQSOMTSGGDDG 483
DB 585 QPPLVGTNTLLSRDSCSSAQAYDTPRKGRGLPAPDELVYOVPOSTQEVSGAGRDG 644

```

RESULT 11

```

US-09-874-198-2
; Sequence 2, Application US/09874198
; Patent No. US20020082208A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.
; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; FILE REFERENCE: 09011-002002
; CURRENT APPLICATION NUMBER: US/09/874,198
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-198-2

```

```

Query Match 6.6%; Score 177.5; DB 10; Length 686;
Best Local Similarity 24.4%; Pred. No. 4.3e-06;
Matches 52; Conservative 39; Mismatches 97; Indels 25; Gaps 7;

```

```

OY 32 CGHIVTSQDSGTMTSKNRYPTNYVCEKIIIVPKGRLLI---RLGDINIESKT-CAS 87
DB 184 CSQOVFTQSGELSSPEYPRPYKLSCTYSISLEGFVILDFVESFDVETHPETLCPY 243
OY 88 DYLLSSATDQYGYCGSNAVPEKRLNSNEVTLF---KSGSHISGRGFLTYASSDHP 144
DB 244 DFLKIQTDRHEHPFCGK-TLPHRIETKSTVTYITVETDESGH---TGKRIHYTSTAOP 299
OY 145 DLITCLERESH-----YEEKYKFCPCAGCDIAGDISGNT-----KQGYRDTSL 190
DB 300 CYPMAAPPNGHVSPOVAKYILKDSFSIFCETGYELLQGHILPKSFTAVCKDKDSMDRMP 359
OY 191 CKAIHAGITDELGHINILQSGISHYEGLL 223
DB 360 ACSIVDCGPPDDLPSGRVEXITGPGVTYTKAVI 392

```

```

RESULT 12
US-09-874-238-2
; Sequence 2, Application US/09874238
; Patent No. US20020082209A1
; GENERAL INFORMATION:
; APPLICANT: Jensenius, Jens Chr.

```

```

; APPLICANT: Thiel, Steffen
; TITLE OF INVENTION: MASP-2 COMPLEMENT-FIXING ENZYME, AND
; FILE REFERENCE: 09011-002003
; CURRENT APPLICATION NUMBER: US/09/874,238
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/054,218
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 60/042,678
; PRIOR FILING DATE: 1997-04-03
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-238-2

```

```

Query Match 6.6%; Score 177.5; DB 10; Length 686;
Best Local Similarity 24.4%; Pred. No. 4.3e-06;
Matches 52; Conservative 39; Mismatches 97; Indels 25; Gaps 7;

```

```

OY 32 CGHIVTSQDSGTMTSKNRYPTNYVCEKIIIVPKGRLLI---RLGDINIESKT-CAS 87
DB 184 CSQOVFTQSGELSSPEYPRPYKLSCTYSISLEGFVILDFVESFDVETHPETLCPY 243
OY 88 DYLLSSATDQYGYCGSNAVPEKRLNSNEVTLF---KSGSHISGRGFLTYASSDHP 144
DB 244 DFLKIQTDRHEHPFCGK-TLPHRIETKSTVTYITVETDESGH---TGKRIHYTSTAOP 299
OY 145 DLITCLERESH-----YEEKYKFCPCAGCDIAGDISGNT-----KQGYRDTSL 190
DB 300 CYPMAAPPNGHVSPOVAKYILKDSFSIFCETGYELLQGHILPKSFTAVCKDKDSMDRMP 359
OY 191 CKAIHAGITDELGHINILQSGISHYEGLL 223
DB 360 ACSIVDCGPPDDLPSGRVEXITGPGVTYTKAVI 392

```

```

RESULT 13
US-09-864-761-45371
; Sequence 45371, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

```

```

: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 45371
: LENGTH: 75
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC013497.4
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
: OTHER INFORMATION: EST_HUMAN HIT: A1565996.1, EVALUE 2.00e-40
: OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUE 1.00e-09
US-09-864-761-45371

Query Match          6.4%; Score 170.5; DB 10; Length 75;
Best Local Similarity 54.8%; Pred. No. 7.2e-07;
Matches 34; Conservative 9; Mismatches 18; Indels 1; Gaps 1;

OY 30 DCGGHTVSDSGTWSNRYGTYPNVYCEKIIITPKGRLLNLGDLNIE-SKTCASD 88
Db 1 DCGGHTVAGPESGTLTSINYQTYPNSTVCEWEIHKMGGERIKFGDIEDSDSCHFN 60

OY 89 YL 90
Db 61 YL 62

RESULT 14
US-09-905-291A-104
: Sequence 104, Application US/09905291A
: GENERAL INFORMATION:
: APPLICANT: Gementech, Inc.
: APPLICANT: Ashkenazi, Avi
: APPLICANT: Botstein, David
: APPLICANT: Desnovers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, A.
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth, J.
: APPLICANT: Kiljavin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
```

```

: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: 10466-14
: CURRENT FILING DATE: 2001-07-12
: PRIOR APPLICATION NUMBER: US/09/905,291A
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 104
: LENGTH: 415
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-905-291A-104

Query Match          6.4%; Score 170.5; DB 9; Length 415;
Best Local Similarity 23.3%; Pred. No. 8e-06;
Matches 99; Conservative 67; Mismatches 152; Indels 107; Gaps 22;

OY 4 GAGGPRVYALAFVACAPLRQ--AEELGDCGHTVSDSGTWSKNTPTGYTYCEK 61
Db 3 GANMAAPCLLLAATQTSRQOSPERPYFTGCGILTG-ESGFISSEGPVYPNPKCTW 61

OY 62 IITVPGKRLILRLGLDNIES-KTCASDYLLFSSATDQY-----GPGCGSAVPE 111
Db 62 KITVPEGRVYVLANRFIDLESNDLCRIDFV-----DYTNHGANGORIGRCGIFR-PGA 114

OY 112 LRLNSNEVTVLFKSGSHISGRFLITYASS-----D 142
Db 115 LVSSGNKMMVMISDAMTAGNGFAMPSAAPNKRGDQYCGGLDRPSSGSKTPNMPDRD 174

OY 143 HPDLITC-----LENGSHYFEKYSKFCPAG---CR-DIAGDISGNTDGRDLSLCKAA 194
Db 175 YPAGVTCWMIYAPKRNQILIEIKFEKFDYERDNYVAVFNGGEVNDARRIGKCGDS 234

OY 195 IHAGIIDELGHNINLQSGISHYEGLLANGVLSRRGSLSEKFFLTTPPMNTTYAIP 254
Db 235 PPAPIVSEK-----NELLIOPLSDL-SLTADGFIQHY--IFRPKLLPTTEQPVTT-TFP 285
```

```

; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO: 104
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-902-853-104

```

```

Query Match          6.4%; Score 170.5; DB 9; Length 415;
Best Local Similarity 23.3%; Pred. No. 8e-06;
Matches 99; Conservative 67; Mismatches 152; Indels 107; Gaps 22

OY 4 GAGGGSVALALFVACAPRLQ--AELBAGCGGHVTSODSGTSTSNNTGTPNTVYCK 61
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 3 GANAMAPLCLLAAATQLSRQOSPSPRPVFTCGGILTG-ESGFISGFGVYPPNSKCTW 61
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 62 ITTVEGKRLIRLDLNIES-KTCASDYLFSSATDQY-----GPGCSMAVPKE 111
   ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 62 KITVEGKVVYVNFPIDESNDLCRYDV-----DYNGHANGORIGRCGCTPR-PCA 114
   ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 112 LRLNSNEVTLFEKSGSHISGRGFLITYASS-----D 142
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 115 LVSSGNKMMVOMISDANTAGNGFMFMFSAEEPNERGDQYCGGLDPRSGSFKTPNMPDRD 174
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 143 HPDLITC-----LERSHYPEEKYSKPCPAG--CR-DIAGDISGNTKDGYPDTSLLCKAA 194
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 175 YPAGVTCVWHIYAPKNQILFEKFDVERDNYCRDIYAVFNGGEVNDARRIKYCGDS 234
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 195 IHAGITDELGSHINLOSRSKISHYEGLLANGVLSRHGSLSEKRELETTPGMNITTVAlP 254
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 235 PPAPIVSE-----NELLIQFLSDL-SLWADFGIGHY-IFRPKRLPTTTEQPVTT-TPP 285
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 255 SVIFALLLTGM-GIFATCRKRRKK-----GNPYVSADAOKTGCKWQIYPPARHSTEF 308
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 286 -----VTTGLKPTVALCQOKCRRTGLEGN-YCSSDVLVAG--TVITTTTRDGSLLHA 334
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 309 TISYDN-----EKEMTOKLILTSPMADYQOPLMIGTGVVARKGSTFFRPMPTD 356
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 335 TVSIINIRKESMLAIQAGKKNMSARLTVCK-----QCPLL-----RGLNIITMGV 382
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 357 TEEVR 361
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 383 GEDGR 387

Search completed: May 15, 2003, 13:21:36
Job time : 20.9319 secs

```

THIS PAGE BLANK (USPTO)